Studies of the *rough* Homeodomain of *Drosophila melanogaster.*

A Thesis Presented for the Degree of Doctor of Philosophy

in the Australian National University

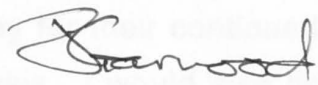
by
Jeremy John Garwood

Research School of Chemistry
Canberra

April 1990

DECLARATION

The work presented in this thesis is the original work of the candidate except where otherwise stated.



Jeremy John Garwood.

ACKNOWLEDGEMENTS

I wish to express my thanks to my two supervisors, Dr Rob Saint and Dr Nick Dixon for their support and guidance during my studies at the Australian National University and CSIRO, Division of Entomology.

My project began at the CSIRO, Division of Entomology and my sincere thanks are due to Drs. Bill Kalionis, Max Whitten, Marion Healy, Robyn Russell, John Oakshott, Ms Hilary Mende, Mr Zung Doan and Mr Eric Hines for their advice and technical assistance throughout this work.

At the Australian National University, I would like to express my gratitude to Dr Phil Hendry, Mr Phil Thompson, Dr Jenny Beck and Ms Penny Lilley for their continued support and assistance throughout this thesis. I would also like to thank Drs. Mike Howell (Zoology), Dennis Shaw, and Subash Vasudevan (JCSMR).

I am as ever endebted to my parents and brothers for their encouragement, support and tolerance during my many years of study, and very much look forward to seeing them again after this long sojourn away from home. To Freda and Harry Buxton for making Christmas in Australia an enjoyable family occasion, Kieran Coyle for providing tea and talk on many pleasant occasions, Arthur Witherall for his idiosyncratic approach to it all, and Heather Nicholson for nourishment.

I especially dedicate this thesis to Eva Manikis for her continued support, tolerance and good humour during the four years of my study in Canberra. I would like to thank her for her patience and time in helping me to cope with all my trials, and in sharing this experience.

As sure as what is most sure

from *St. Winefred's Well*

As sure as what is most sure, sure as that spring primroses
Shall new-dapple next year, sure as tomorrow morning,
Amongst come-back-again things, things with a revival,
things with a recovery.

Gerard Manley Hopkins (1844-99).

Abstract

The homeobox is a highly conserved region of genes whose protein products appear to have a gene regulatory function. It has been proposed that the homeobox encodes a DNA-binding domain within these proteins (the homeodomain), with structural similarities to well-characterised prokaryotic DNA-binding proteins. The *rough* gene of *Drosophila melanogater* has a role in regulating the differentiation of photoreceptor cells in the developing ommatidium of the eye-imaginal disc. It has been shown to encode a homeodomain with which it might interact specifically with DNA to regulate gene expression.

In this thesis are described studies of the expression, structure and function of the *rough* homeodomain. The *rough* homeobox was manipulated into a bacterial gene expression vector such that it encoded a 70 amino acid polypeptide (containing the entire homeobox). This gene construct, pJG(2), was successfully expressed, and the *rough* homeodomain (*ro*HD) protein was purified as a stable entity. The purification was hampered in the absence of an assay for the protein and insufficient quantities were obtained to permit structural studies. However, the purified *ro*HD was shown to possess a sequence-specific DNA-binding activity, showing affinity for two previously characterised homeodomain consensus sequences, $(TAA)_5$ and TCAATTAAAT ($K_d \sim 10^{-8}$M). Preliminary evidence on binding to sequences in the promoter region of *rough* is also presented. A comparison of the *ro*HD sequence with that of the *Antennapedia* homeodomain, whose 3-D structure has recently been determined, suggests that they are very similar.

A fusion protein containing most of the *rough* homeodomain was produced (X17). Polyclonal antibodies were raised against this

tissue are presented. These suggest that the *rough* protein is expressed in cells just posterior to the morphogenetic furrow in eye-imaginal discs. A protein ~ 40 kdal is evident on Western blots of third instar larval protein extracts which may correspond to the *rough* protein.

Table of Contents

CHAPTER 1.

General Introduction and the Scope of the Thesis.

1.1 The Homeobox and Homeotic Mutations

The homeobox was first discovered as a result of sequence analysis of genes whose mutations resulted in homeotic transformations in *Drosophila melanogaster*. Such homeotic mutations, first named as long ago as 1894 (Bateson 1894) cause developmental anomalies in which one part of the body develops in the likeness of another. The possible significance of these abnormalities for the study of biological development was especially noted by Bateson, given the occurrence of such altered developmental pathways in many plants and animals.

The genetic induction of homeotic transformations has been successfully used to identify many genes that control growth and pattern formation during development. Homeosis has been more extensively studied in the fruit fly, *Drosophila melanogaster*, than in any other organism. This has been aided by the popularity of *Drosophila* as an experimental organism for geneticists since the turn of the century and by its rapid and accessible development. Because of its experimental advantages, and given the lack of understanding about the genetic control of development in any organism, much effort has been devoted to the study of the development of this insect in the hope that it will provide insights into the development of other animals, including mammals, which are not so amenable to genetic analysis.

In *Drosophila*, homeotic transformations can be caused by mutations in any of at least a few dozen genes. Most research to date has focussed on two clusters of homeotic genes: the bithorax complex (Lewis 1978) and the antennapedia (Antp) complex

(Kaufmann et al. 1980). Each of the genes in the two complexes is expressed in a certain region of the *Drosophila* embryo during early development and, in the absence of a particular gene's function, the corresponding part of the embryo follows a different developmental pathway (Akam 1987). Because these genes appeared to have similar functions but in different places, it was suggested that they might all have evolved from an ancestral gene by duplication and divergence, and that such divergence of the homeotic genes could have led to the evolution of insects, with their varied body segments, from annelid-like ancestors with simpler repeating patterns of body segments (Lewis 1978).

The prediction that the genes of the bithorax and antennapedia complexes would have related structures was indeed shown by determination of the nucleotide sequence of the *Ultrabithorax* and *Antennapedia* genes. This revealed that they both contained a similar sequence of about 180 base pairs which was named the homeobox (McGinnis et al. 1984a,1984b). The DNA sequences of these two homeoboxes are 74% identical, whilst at the protein-coding level they are 88% homologous (54 amino acids out of 61). This observation of greater conservation of the protein rather than the DNA sequence suggests that it is the protein sequence which is being selected and maintained during evolution (Scott et al.1989).

Homeoboxes have now been found in over 20 *Drosophila* genes, most of which are known to regulate development. They include genes which control segment differentiation (such as *Ultrabithorax* and *Antennapedia*), segmentation genes that control the division of the embryo into segments and that in some cases control the homeotic genes (like *fushi tarazu* and *engrailed*), genes involved in dorsal/ventral differentiation (*zen 1* and *2*), genes that act maternally to control anterior-posterior polarity of the embryo (*bicoid* and *caudal*), a gene (*cut*) that controls cell determination in

the peripheral nervous system and also functions in leg, head and wing development, and the *rough* gene which functions in eye development (Scott et al.1989). Some of these genes were isolated using homeobox DNA cross-hybridisation, whilst in other cases the homeobox was only discovered after the gene had been identified in other ways.

In addition, more that 50 homeobox sequences have been obtained from other organisms. Homeoboxes quite closely related to fly homeoboxes have been found by DNA cross-hybridisation in mice, humans, chicken and *Xenopus* (McGinnis et al.1984b) and in representative nematodes, annelids, arthropods, ascidians, echinoderms, brachiopods, tapeworms, molluscs and chordates, but not in bacteria, fungi, sponges, slime moulds, flatworms, platyhelminths, or aschelminths (Scott et al.1989). In most cases, the functions of these genes are unknown but a role in transcriptional regulation has been proposed based on the evidence derived from studies of the *Drosophila* homeobox genes. The evidence for transcriptional regulation is, briefly, that all of the proteins encoded by homeobox genes thus far have been found in the cell nucleus; that there is a similarity in structure between the homeodomain (the polypeptide encoded by the homeobox) and known prokaryotic DNA-binding proteins; and finally, that the homeodomain binds specifically to certain DNA sequences *in vitro*.

The review by Scott et al. (1989) of the structure and function of the homeodomain discusses this, and presents a comparison of the homeodomain sequences, a model of the homeodomian structure, and the DNA-binding studies.

The distribution of homeodomain proteins is discussed in this thesis in Chapter 5, whilst the structural aspects are presented in Chapter 6. A discussion of the DNA-binding studies is given in Chapter 4.

1.2 The *rough* Gene and its Phenotype.

The *rough* mutation in *Drosophila* was first discovered by Muller in 1913 (Tomlinson et al.1988) and derives its name from the roughened appearance of the normally precise hexagonal array of ommatidia in the compound eyes of the mutant fly.

Two groups have independently cloned and sequenced parts of the *rough* gene, almost simultaneously discovering that it contains a homeobox.

The first group (Saint et al.1988) isolated the *rough* genomic fragment containing the homeobox from a *Drosophila* Canton-S chromosomal library by means of a DNA cross-hybridisation experiment with a 33-base oligodeoxynucleotide homologous to a consensus homeobox sequence. This fragment was sequenced extensively to show that the homeobox lay in an open reading frame of 333 base pairs (as shown in Figure 2.4a)). It was this fragment that was subsequently used for all of the expression work presented in this thesis. *In situ* DNA-mRNA hybridisation studies localised the expression of the *rough* gene to cells within, and posterior to, the morphogenetic furrow (the site of the primary pattern-forming events) in the developing retina, and also to a region of the brain in the third instar larva.

Meanwhile, the other group (Tomlinson et al.1988a) succeeded in obtaining the *rough* gene by isolating the genomic region corresponding to a mutant *rough* allele. The equivalent genomic region from a Canton-S wild-type library was then isolated. Restriction fragments of the *rough* gene were subsequently used to screen an eye-imaginal disc cDNA library (that is DNA sequences derived from the messenger RNA present in this tissue), and two cDNAs for the *rough* gene were obtained and sequenced. It emerged that the homeobox lies in the second of the

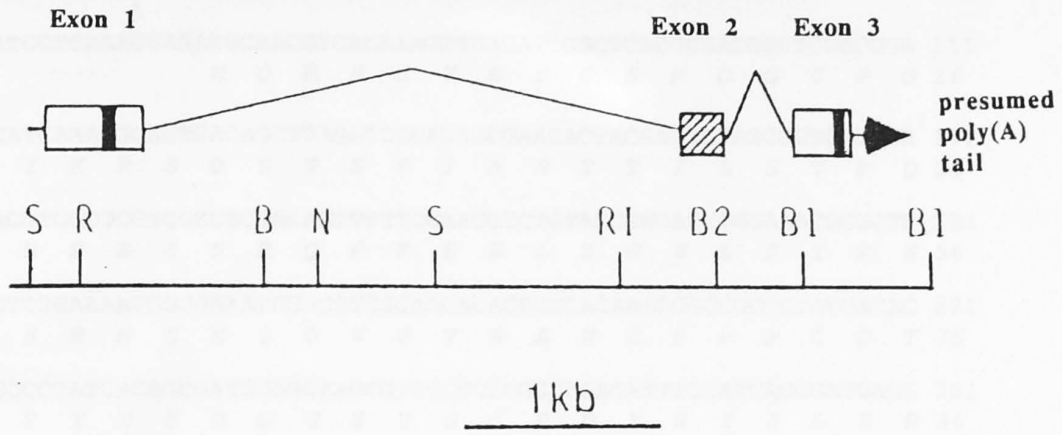


Figure 1.1a): The *rough* transcription unit (Tomlinson et al.1988)

The *rough* transcription unit spans ~4.3 kb and consists of 3 exons and 2 introns. The homeobox is in exon 2 (cross-hatched box). The solid bars in exons 1 and 3 indicate *opa*-like repeats (residues 116-138, and 322-338; discussed in section 4.1.4). The restriction enzyme sites shown are: S, *Sal* I (-53,2063); R, *Rsa* I (209); B, *Bam* HI (1170); N, *Not* I (1448); RI, *Eco* RI (3028); B2, *Bgl* II (3549); BI, *Bal* I (3993,4666). Distances are from the start of transcription.

Figure 1.1b): Nucleotide sequence of the *rough* gene (Tomlinson et al.1988).

Nucleotide 1 marks the transcription start site. The exon-intron boundaries are underlined. The conceptual translation of the 350 amino acid *rough* protein is given in the one-letter amino acid code below codons.The homeodomain is underlined in exon 2 (Saint et al.(1988) show the homeodomain to start with Arg-190, and finish at Glu-250, whilst Tomlinson et al. show it starting at Gln-191, and finishing at Lys-251).

EXON 1

INTRON 1

```
GTGGCACTTCAGTCGCTCGCCGATTACCCAAGTGATTAGTTTGGACAGCTAACGCGGCAC 51
          1
ATCCTCAAACGAAATGCAACGTCACAAAGTTGAGATTGGCTCACCCGACGGCTCGCCGGG 111
             M  Q  R  H  K  V  E  I  G  S  P  D  G  S  P  G  16

CATCAAAAGGAGTGACAGCTTAGATCCCATAGCGAACACTACGATTCTATCCGTGCCGCA 171
  I  K  R  S  D  S  T  D  P  I  A  N  T  T  I  L  S  V  P  Q  36

ACGTCCCTCGTCGCCGCGACAGTTTTTCGAACGTCTGTACGGGCATTTGGAAACGCGCTC 231
  R  P  S  S  P  R  Q  F  F  E  R  L  Y  G  H  L  E  T  R  S  56

CTCGGAAAATGGGGAAATCGACGTCGGAACACACGCCCACAAGCCGCCGCCCTGTGACAC 291
  S  E  N  G  E  I  D  V  G  T  H  A  H  K  P  P  P  C  D  T  76

GCCCTATCACAGCGATGGAGGAAGCGTCTCCCTCGCCAGACATTTCCATCAGCGATGAGC 351
  P  Y  H  S  D  G  G  S  V  S  S  P  D  I  S  I  S  D  E  R  96

GCACCTCGCTAGCCGCATATCCCGCCTACGATTTCTATGGTCACGCCAAGGACTATCCCC 411
  T  S  L  A  A  Y  P  A  Y  D  F  Y  G  H  A  K  D  Y  P  Q  116

AGCATCCTAGCCAGCAGCACCAGCAGCATCATCATCATCACCATCATCCGCCGCAGCTGG 471
  H  P  S  D  Q  H  Q  Q  H  H  H  H  H  H  H  P  P  Q  L  V  136

TGCACCAGAAACTCAGCTACGTGAGCCCGCCGCCAGCAATTGCCGCCGGTGGAGCTGCGA 531
  H  Q  K  L  S  Y  V  S  P  P  P  A  I  A  A  G  G  A  A  N  156

ATCCTGTCCTGCCACACGCATTCCCAGCCGGCTTCCCAAGTGACCCGCACTTCAGCGCCG 591
  P  V  L  P  H  A  F  P  A  G  F  P  S  D  P  H  F  S  A  G  176

GCTTTTCCGCATTTCGTGAGTACTTACATTTAAAAAAGAATGGTTTCTAACAATGTGTAG 651
  F  S  A  F  L                                                181

AATATTTGCAAAGCTATAGAGAATCTTTAGAACTACGAAAATAACTTCTTCAGAACCTTA 711
ACTCTTACATGTTTATTGTGAGGTTTCTTAGTTCTTACACACTAAGTACAAGAGTAAATT 771
TGAGCTACTACGAATTTGGGTCCTAACTAGATATTTAGGTGAGCGGGGATTTGTACTACG 831
GCGGGGGCGAGGATTTCTTTGGCTTTCCCTGCTTCCTGCTTTCTCCGCCGCCGCCTTCTC 891
CTCCCGCTGCTTAGCCATCTGGCGCAGCTCCAGGCTGAGGAACATCTTCAGGTTCGGGTC 951
GGTGATCTTGCTCAGCACAATCTCTCGCGCCTCCTCATCGTTGTCGCCCAGCAGCTTGGC 1011
CAGCGTCTGTGCGTAGTCCAGCCAAACGCAGTTGGCGCAGCCCGACATGCAGCAGGTGGT 1071
GGGCTCGGGAGGTATCTGCAAACGATTATATAATATGCAATCTCTTGAGCCACTTTCCTC 1131
TTAGGCTACAAAGTTATTTCAGTATTGTGATAACAATTGGATCCTTCAATAATACTGATC 1191
AATAGTTTTAGAAGTTCATAGAACTACACTAAGTTTTGTGTTACAATACTAGATCGATGA 1251
TCCTTTCTTTCTAATGAACTACAATCTTACCAACTGATATTTGACCTATAATAGCAATGA 1311
TAACAACTAGCTTACAACATTTACCTCTATGTTTCTGAGCCGCTTTCTTCCCTTGGTGGT 1371
TTTATCTTTGGCTTTGGGAGTGTCAGTGCCCGTCGTTGTGCTATTGGGCGGACAGCCACT 1431
TGCTGCCTCCTGGGCGGCCGCATCCTTGCTGCTGGCATCATCCTGGGAGCTGCTGCCCGA 1491
CAATTGGCGGGCCAGTGAGGCCCAACGCTCGGCGCCACGTCGCAGAAGCAAGCATCGCGG 1551
ACAGCTGAAGCCCAGGGGAGCCAACATTTCCGCCCGATCCCCGAAAGAGTTCGAAAGTTA 1611
CGATTGAAAAGGCCCAGAGGAAGTGATGAGCATTTCACACGCCTGCGCAGGTGCAGCCCA 1671
GATGAACTTGGTTCCGCAGGAACTGCAAAAAAAAACTGATTAATCGTCAGTGCTGCCATA 1731
TTGAATATCCCCCACATCTTCGCACCGAGTGCGCTGGCCAGGATTATGTAAAATCAGTTG 1791
GGAAAAAATAAACGAAACTAAGGTGGGGCAACCAAGCAATGTCCACACAATGAGCTGCTG 1851
CTGCTGCTGCTGCTGTGCTGCTGCTGACCCCCAAATGCAAAGAAATTAGCCAAGGAAAAA 1911
AAATAACCGAATGCCAAGAAAAAATTAAAACGCATTGAGCCGTGAAAGTTTTCGTTTACT 1971
TGCGGCTTTTTCGTTTGCTCCTGATGAGGTGGTCCAAGGTGGCCATGTGCGGAGATGAGG 2031
TTGAGGGGATATTGAAACATGGAAATGCGTCGACTAATGAAAATGATCGCAGGTTAAGGG 2091
GTGAATGGGTTAGCCATTCGTAAGCGAGAAATCAAAGTTAAAGGCACTTTGAAACATTTT 2151
CTTCGAGCAGAAACGTATTCTACATATGCAAGCATTTTTTAGATTCAGAACAGTTGATAT 2211
ATTCAATGCTAAACCAGCTGGATAGTTACTATAAGTACTTGATCTTTAAATTAACTAACT 2271
TACAAGCCAAACTTTTGGTCAATCTAAAAAAAATAACAAGTGCCTGAAAAAATTAGCATTG 2331
```

```
TAATTGAGCTCGGCAGGAAACCTAAAAAAGCAAACACAAACAGCAAGCGAGGAGGGTTAA 2391
GCGAGGAAGACTCACTTTTAGCCACAACAAACCGTCCGAAAGGACACTTGGCTATCGAAT 2451
TTCACACGTGATTTAAAAAAAAAAAAACCTTCGGCAAGGATCAGGGAAGAGCTTAAGCCC 2511
CTTCAGAGCCATCCGGACTAGGGGTAATCAAGCAAAAGCTAGAGATTACAACGCCCACAA 2571
AGCAACCCCTTCCCGGGGAAACAACCCTCCCCCCCCCCACCACCTAACACCACCGAGCAA 2631
CTTGTCAGCTTCGCCTTGTTACATCAATTAGACTCTAAACCGCACCATTCACAATGGAGT 2691
GGGTGTTTGTGGTGGGTGTTCCTGGGTGGCAGGTGCTCGTTGTTGCAGGATGAAGTGGGG 2751
CATAATGTGCGCTTGGTTGAGCTTCATCGTCGGCGACACTGCGAATGCAAACAATCAAAC 2811
GACAGAATGGAAAACTCGAAAGAGCGGATGATGCCGAGAATTTCGACGAGGATGATGATG 2871
ATGATGATGATGGCCAGCCAGCTGTTGGCATTGTTATAAATTTCCCGGATCAGCATAGAT 2931
TGCCGAAATGAAAAGCGTGGTAAAAAAAAAAACATAGAGAACAGTAACTACGTGGTAAATG 2991
CACTTCTAATGTGCACCACAAACGCGACCCACGACGAATTCTGGCAGGTGGGCGGGCGGA 3051
GAGAGCGAGAGAGGGAGGACGACGGACGGACCACGGATCGGAGTGGTTAATTATTTGTTT 3111
AACTGACGCAGCAACACGGCAACCACATTCGAAAACGTTTAGTCCCGCTCATAACTTGAT 3171
GGTTTGTTAATGTAAGATTAATTTCGCTCTAACATCTCCATTGTGTTGGACATAAACGAA 3231
AGTGGGTGGGCTGTGAAAAAATGATGCTATCCGACGGATGTCGTTATTGTACACACTGAA 3291
TGGTGCGGGGTTTACATAACTAATAAGCTGATTCCAACTAGATATTATATACGTATTAAT 3351

TGTAAGGATATCAATGTATTTGACTTTCAGTGGCTCGCAGGCGACGCAAGGAGGGAAGAC 3411
                            A  R  R  R  R  K  E  G  R   190

AACGCCGCCAGAGGACCACTTTCAGCACAGAGCAGACGCTTCGCCTGGAGGTGGAGTTCC 3471
Q  R  R  Q  R  T  T  F  S  T  E  Q  T  L  R  L  E  V  E  F   210

ATCGGAACGAGTACATCTCCAGGAGTCGTCGCTTCGAGCTGGCCGAAACGCTGCGTCTGA 3531
H  R  N  E  Y  I  S  R  S  R  R  F  E  L  A  E  T  L  R  L   230

CCGAAACGCAAATCAAGATCTGGTTCCAGAATCGCAGGGCCAAGGACAAACGCATTGAAA 3591
T  E  T  Q  I  K  I  W  F  Q  N  R  R  A  K  D  K  R  I  E   250

AGGCTCAGATCGATCAGCACTACAGGTGGGTTTCCCCTTTAGCTAATTTGAATCGTTTTT 3651
K  A  Q  I  D  Q  H  Y  R                                     259

TCTGTTTCATATTTACAAAGATATCTTCGATTTTGATTGAGTTAAATAACATTAAATGAA 3711
CATAAATTAAAGGTTGATGAATAGCTCGTGCAGAGGATTTTAACTGGTTATTTATACCTG 3771
CATTTTCGAAACGGAAGTAATTTAGTGAAATGTCACTAGTTTAATTGATGGCCAACATAA 3831
AATTAACACGAAATTAATTTTACGACATTTATGCGCACACAGACCCAATTAGACGATATC 3891
ATTTTTTTCCATATTTTTGTTTTTTTTTTTTTGTGCCGGAGGGGGTTGCTCGAAATTATT 3951

TTGTACATTGATGTAATCAATTTTTCAGAAACTTTGTCGTGGCCAATGGTTTTATGAGCT 4011
                                N  F  V  V  A  N  G  F  M  S   269

CTATAATGGGTCAGGCTGCCACGACCATGCCCCCCGGAGGCGTCACTGGGGGCGTGGCGG 4071
S  I  M  G  Q  A  A  T  T  M  P  P  G  G  V  T  G  G  V  A   289

TGGGCGTGGGCCTAAACTACTATGCCGCGGCAGCGACGCCGACAGCGCTGCCCAAGGATA 4131
V  G  V  G  L  N  Y  Y  A  A  A  A  I  P  T  A  L  P  K  D   309

ACACGCAGGATGCCAATTTCATCGACATCGATGACCAGTTCCAGCGCCAGCAACAACAAA 4191
N  T  Q  D  A  N  F  I  D  I  D  D  Q  F  Q  R  Q  Q  Q  Q   329

AACAACAACAACAGCAGCAGCAGCAGCGACGTCGAGAAACAACAACGCCTATTAACATTT 4251
K  Q  Q  Q  Q  Q  Q  Q  Q  R  R  R  E  T  T  T  P  I  N  I   349

GCTAATACTT 4261
C  *        350
```

EXON 2

INTRON 2

EXON 3

Figure 1.1 b): Nucleotide sequence of the *rough* gene Details in Figure 1.1 a).

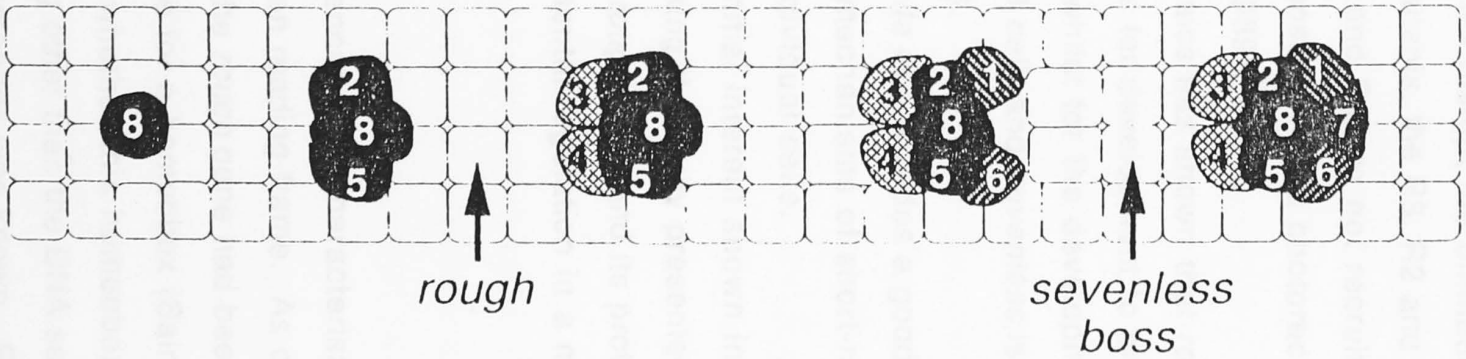
three exons that comprise the gene, which together give a 1.3 kb mRNA, encoding a predicted protein of 350 amino acids. The organisation and sequence of the *rough* transcription unit is shown in Figure 1.1.

Analysis of *rough* mutations shows that they disrupt an early stage in the ommatidial assembly of the developing eye imaginal disc (Tomlinson et al.1988a). Figure 1.2 presents a simplified summary of the stage at which the *rough* mutant results in the malformation of the ommatidia.

The mature retina of the compound eye is made up on the surface of an array of approximately 800 twenty-cell units, called ommatidia. Each of the 20 cells in an ommatidium can be identified by its morphology and distinct position. The differentiation of the cells that comprise the ommatidium begins in the third instar larva in an initially unpatterned monolayer of epithelial cells called the eye imaginal disc. Ommatidial assembly does not occur synchronously throughout the disc but instead begins at the posterior edge and progresses anteriorly (Tomlinson et al.1988b). The first eight cells in each ommatidium to differentiate are the photoreceptor cells. Differentiation of these occurs in a fixed sequence, beginning with central R8 photoreceptor and proceeding pairwise with R2 and R5, followed by R3 and R4, then R1 and R6, and finally, R7. It has been established that no lineage relationships exist between the photoreceptors cells, which indicates that their fate is determined by their environment. A model to explain these observations has been presented (Tomlinson and Ready 1987). This proposes a sequential series of cell inductions. Thus, as the R8 cell begins to differentiate, it expresses a signal which instructs only two of the cells in its immediate vicinity to become the R2 and R5 photoreceptors. Subsequently, the R8, R2 and R5 cells provide a signal which recruits two more cells to become R3 and R4.

Figure 1.2: Differentiation sequence of photoreceptors in the ommatidium of *Drosophila melanogaster* (from Rubin (1989)).

The individual stages displayed differ by approximately 3-4 hours of developmental time. The *rough* mutant causes ommatidial assembly to go wrong between the onset of differentiation of the R2/R5 and R3/R4 cells, whereas *sevenless* and *boss* cause a breakdown between the onset of differentiation of the R1/R6 and R7 cells.



Similarly, the recruitment of the last three photoreceptors, R1,R6, and R7, would occur with two additional cycles of cell induction.

Studies of the *rough* gene, and of the *sevenless* (*sev*) and *bride of sevenless* (*boss*) genes, tend to support this model. Mutations in each of these genes disrupt the ommatidial assembly at a discrete point: in *rough* mutants, the R8, R2 and R5 cells differentiate normally, but R3 and R4 are not recruited correctly; whilst in both *sevenless* and *boss*, the final photoreceptor, R7, fails to differentiate (Rubin 1989).

Analysis of somatic mosaics has shown that *rough* is required in the R2 and R5 cells for development to proceed normally to the R3/R4 stage, whilst for the development of the R7 cell, *boss* is required in the R8 cell, and *sevenless* is required in the R7 cell itself (Rubin 1989).

The developing *Drosophila* eye provides a good system in which to study the molecular mechanisms of short-range inductive interactions at the level of individual cells.

Thus, in addition to the initial interest shown in it as a homeobox gene when undertaking the study presented in this thesis, further studies of the *rough* gene and its protein may provide insights into developmental regulation in a model system.

1.3 Scope of Thesis

The current work commenced with characterisation of a novel homeobox-containing open reading frame. As described above, the homebox region of the *rough* gene had been isolated from a genomic library using a probe for a homeobox (Saint et al.1988). At the time, it was not known whether this homeobox was a part of any known gene locus. Nothing other than the DNA sequence of the 3.0 kb region containing the homeobox was known. From this sequence, it appeared that the homeobox lay within an open reading

frame of 333 base pairs encoding a maximum of 111 amino acids and delineated by two TGA stop codons.

Although attempts were being made to isolate the cDNA for this gene (Saint et al.1988), it was decided to concentrate this project upon the expression of the open reading frame bearing the homeobox region.

This work was anticipated to develop in three main areas and, with differing emphasis, has developed as such.

a.The first aspect of the project entailed using an approach which had been successfully utilised with other homeobox genes: the generation of polyclonal antibodies against a fusion protein bearing the *rough* homeodomain (*ro*HD), and, in turn, carrying out immunohistochemical studies upon the life cycle of the insect to gain temporal and spatial information regarding the localised expression of the native *Drosophila* protein. The fusion protein was generated in a bacterial expression system (the pEX plasmid vectors, and *E. coli*). It was partially purified and was used to raise polyclonal antibodies in rabbits. These antibodies were used in a series of experiments involving Western blotting, immunoprecipitation, and immunohistochemistry. These experiments are described in Chapter 5.

b.The second aspect of the project involved expression of the homeobox region. The expression of any homeobox to give a polypeptide of a little larger that 60 amino acids had not previously been achieved. On the assumption that the resulting homeodomain polypeptide would fold stably to give a three-dimensional structure akin to that adopted by it in the complete protein, it was hoped that the developing technique of two-dimensional nuclear magnetic resonance spectroscopy (2-D NMR), might be used to confirm its predicted structure (Scott et al.1989). The current limits to the size of a polypeptide from which useful spatial information may be derived by 2-D NMR is estimated to be

around 90 amino acids, so the proposed homeodomain would be of an appropriate size. Indeed, such an approach might present the first structural information available about any homeodomain, and might confirm the existence of the proposed α-helix-β-turn-α-helix-DNA-binding motif.

In Chapter 2 are described experiments which led to the successful manipulation of the *rough* homeobox-encoding DNA, placing it into a plasmid vector under the control of a strong viral promoter of transcription and bacterial translation initiation signals. In Chapter 3, the successful expression of the *rough* homeodomain is described, together with strategies used to improve the production and purification of the protein.

c.The final facet of the research described herein concerns the functional properties of the purified homeodomain polypeptide. The homeodomain has been proposed to include a DNA-binding motif similar to that in other proteins that act in developmental switches (Scott et al.1989). Although there may be many aspects to the timing and specificity of the genetic regulatory role that these proteins seem to display, the variable specificity and strength of binding of the homeodomain motif to DNA regions found at appropriate sites in the genome of the fly (especially in the promoter regions of genes) has been of especial interest. At the time this study was initiated, there had been some *in vitro* DNA-binding studies reported using fusion proteins. Since then, there have been a number of accounts of work with entire proteins and with the isolated *Antennapedia* homeodomain. These have elucidated a number of binding sites present in the promoter regions of genes, and have given some indication of relative affinities of binding.

In Chapter 4, the DNA-binding studies with the *rough* homeodomain and fusion proteins are discussed and considered in comparison with other such studies.

CHAPTER 2.

Construction of the Homeobox Expression Vector.

2.1 Introduction

In this chapter are described the manipulations of the homeobox open reading frame that led to its correct insertion into a plasmid vector intended to optimise its expression. The vector system used was based on pCE30, a plasmid vector developed for the express purpose of maximising the expression of cloned *E. coli* genes (Elvin et al.1990).

The ways in which this system seeks to achieve this goal are first considered below. In all living organisms, there are mechanisms which regulate and control the degree to which any gene is expressed at any one time. Such regulatory controls are essential to the healthy function of a living cell. The expression of cloned genes to produce high levels of specific proteins is not necessarily bound by such niceties, so all of the regulatory controls may be optimised despite deleterious effects on the host bacterium (Figure 2.1).

The first of these regulatory controls is gene dosage. Most genes are found to occur just once in their native genome, i.e. there is just one gene copy per cell. The pCE30 plasmid, on the other hand, is derived from a high-copy-number plasmid vector, pUC9 (Vieira and Messing 1982). Its derivatives should be present at around 500 copies per cell (Chambers et al.1988), and with it a correspondingly high number of copies of any gene that it bears.

The next factor to consider is the efficiency of transcription of the cloned gene (Figure 2.1). This seems mostly to be regulated by the efficiency with which RNA polymerase binds to and initiates transcription from the promoter region of the gene. This can be influenced by the presence of promoter-specific repressor (and

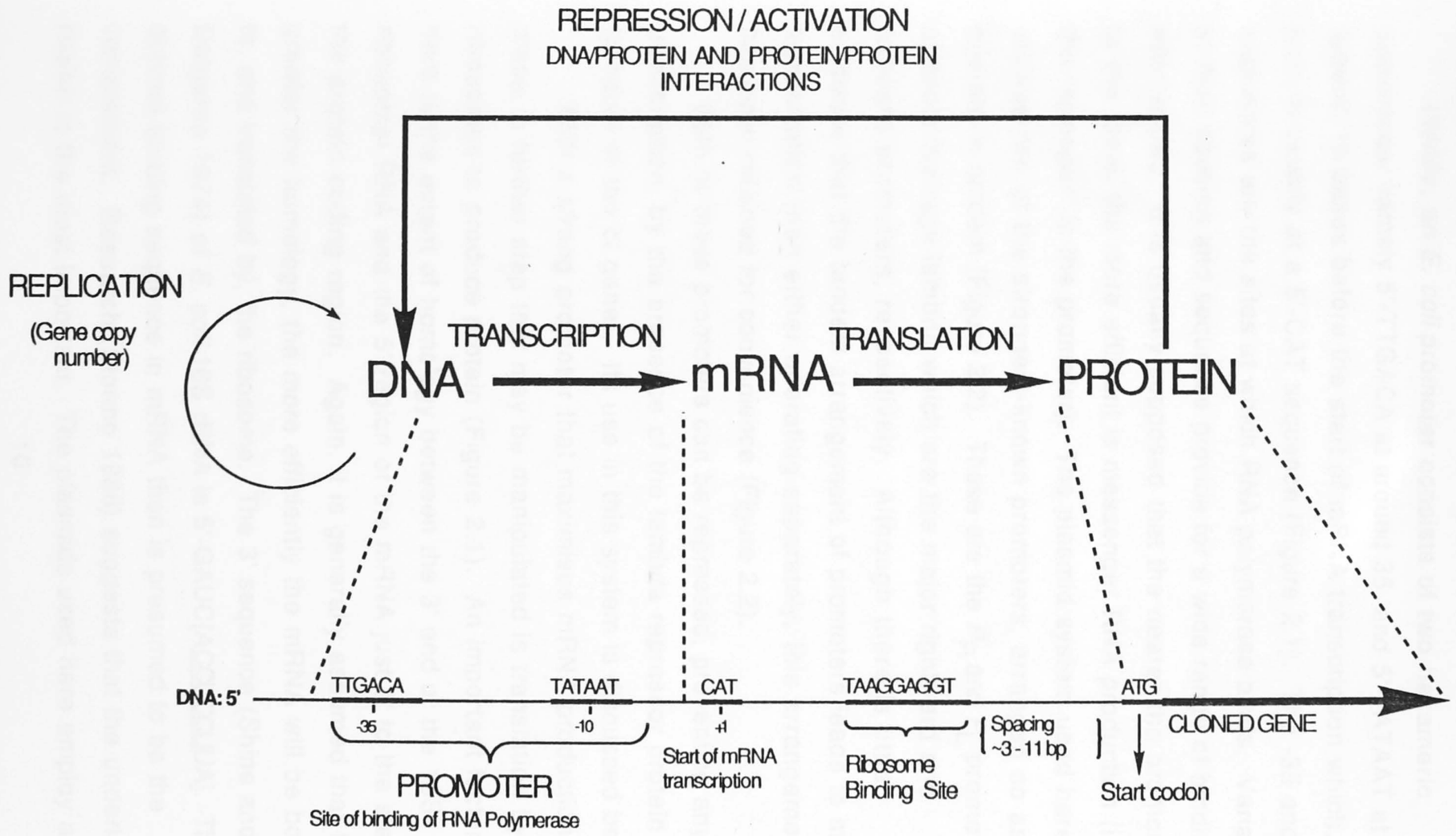


Figure 2.1 ; Levels of control of gene expression in *E.coli*.

enhancer) proteins.

Ideally, an *E. coli* promoter consists of two hexameric sequences, namely 5'-TTGACA at around 35, and 5'-TATAAT at around 10 bases before the start of mRNA transcription which, in turn, is usually at a 5'-CAT sequence (Figure 2.1). The -35 and -10 sequences are the sites at which RNA polymerase binds. Variations in their spacing and sequence provide for a wide range of binding efficiencies. It is usually supposed that the nearer the promoter is to the ideal, the more efficient is messenger RNA production (i.e. the "stronger" is the promoter). The plasmid system used here utilises two of the strongest-known promoters, arranged so as to operate in tandem (Figure 2.2). These are the P_R and P_L promoters of bacteriophage lambda which are the major rightward and leftward promoters, respectively. Although there is little evidence that the tandem arrangement of promoters leads to more transcription than either operating separately, this arrangement has been retained for convenience (Figure 2.2).

Both of these promoters can be repressed, preventing any transcription, by the presence of the lambda repressor protein (product of the cI gene). Its use in this system is discussed below.

With a strong promoter that maximises mRNA production in place, a further step that may be manipulated is translation by ribosomes to produce protein (Figure 2.1). An important factor here is the extent of homology between the 3' end of the 16S-ribosomal RNA and the 5' region of the mRNA just 5' to the start of the protein coding region. Again, it is generally assumed that the greater the homology, the more efficiently the mRNA will be bound to, and translated by, the ribosome. The 3' sequence (Shine and Dalgarno 1974) of *E. coli* 16S rRNA is 5'-GAUC[A<u>CCUCCU</u>UA]. The optimal binding sequence in mRNA then is presumed to be the complement. Research (Stormo 1986) suggests that the underlined region is the most important. The plasmids used here employ a

P_R -35

GGGATAAATA TCTAACACCG TGCGTG**TTGA CT**ATTTTACC TCTGGCGGT**G**

P_R -10 +1 O_{R1}

ATAATGGTTG **CAT**GTACTAA GGAGGTTGTA TGGAACAACG CATAACCCTG

AAAGATTATG CAATGCGCTT TGGGCAAACC AAGACAGCTA AAGATCTCTC

ACCTACCAAA CAATGCCCCC CTGCAAAAAA TAAATTCATA TAAAAAACAT

P_L -35

ACAGATAACC ATCTGCGGTG ATAAATTATC TCTGGCGGTG **TTGACA**TAAA

P_L -10 +1

TACCACTGGC GGT**GATACT**G AGCA**CAT**CAG CAGGACGCAC TGACCACCAT

O_{L1}

GAAGGTGACG CTCTTAAAAA TTAAGCCCTG AAGAAGGGCA GCATTCAAAG

BamHI R.B.S.

CAGAAGGCTT TGGGGTGTGT GATACGAAAC GAAGCATTG**G GATCCTAAGG**

< 9 b.p. > Nco I >> *dnaA* gene >>

AGGTTTAAGA T**CCATG***GTGT CACTTTCGCT TTGGCAGCAG TGTCTTGCCC*

EcoRI

GATTGCAGGA TGAGTTACCA GCCACAGAAT TCAGTATGTG >>

Figure 2.2 : Nucleotide sequence of the Promoter region and start of *dnaA* in pPT148.

The P_R and P_L promoter regions of the λ bacteriophage are shown. The -35, -10, and transcription start sites (+1) are highlighted. The operator sites, O_{R1} and O_{L1}, are also shown; these are the sites to which the λ repressor protein binds when blocking transcription from the promoters. The ribosome binding site (R.B.S.), the ATG start codon, and the start of the *dnaA* gene sequence are shown, including the *Bam* HI, *Nco*I, and *Eco* RI sites utilised in the manipulation of the 5' end of the homeobox described in Section 2.2.1. The location of various sites in pPT148, numbering from the deleted *Eco* RI site in pCE30 (see Figure 2.5) are: P_R +1 (3663), P_L +1 (3876), *dnaA* start codon (4015), and the *Eco* RI site in *dnaA* (4080).

sequence 5'-UAAGGAGGU to the bracketed region above (Figure 2.2).

Other observations suggest that the best initiation codon is AUG (GUG is an alternative), that the optimal spacing between the ribosome-binding site (RBS) and this codon should between 5 and 11 bases, and that this region should preferably be rich in A and U (Stormo 1986). Variations in the spacing and base composition again seem to result in variation in the efficiency of mRNA translation (Stormo 1986). Translation of the vector used here gives an AUG start codon and a spacing of 9 bp. Manipulations used in its construction have allowed 1 G and 2 C's to be included in the spacer region (Figure 2.2).

Another consideration is the stability of the mRNA. Fortunately, it appears that the more the mRNA is translated, the more it is protected from ribonucleolytic digestion, the ribosome effectively blocking its degradation (Kennel 1986). Thus, optimisation of the RBS region probably enhances the stability of the mRNA.

A problem which may now arise, given the strength of gene expression, is the detrimental effects upon the host cell. The production of the protein corresponding to the cloned gene may severely impair the normal function of the bacterial cell to the point of compromising its viability. If this happens from the moment the expression plasmid is introduced into the bacterium then it may prove difficult to generate enough bacterial cell mass from which to extract the expressed protein. The solution is to use an inducible system whereby expression may be controlled by some switch, applied at a stage of growth when there is sufficient cell mass to promise reasonable yields of the expressed gene product. The point at which the expression may best be regulated is at initiation of transcription. This requires the productive binding of RNA polymerase to the promoter region. The binding to the promoter region of another protein, a repressor, may effectively

block the binding of RNA polymerase, preventing any transcription. Repressor proteins are very specific in their recognition of promoters; the *lac* repressor for example binds only to the *lac* operator in proximity to the *lac* promoter. Repressor binding is also reversible. On induction, for example, by addition of an inducer, the repressor will quickly dissociate from the operator, allowing transcription to proceed.

The λ P_R and P_L promoters used in the pCE30 expression system are both subject to repression by the λ repressor protein, the product of the phage *cI* gene. Studies of control of transcription for these promoters resulted in the isolation of a temperature-sensitive allele of *cI*, *cI857* (Sussman and Jacob 1962; Buell and Panayotatos 1986). At temperatures around 30°C the mutant repressor efficiently binds to the λ operators (Figure 2.2), repressing transription from P_R and P_L. On raising the temperature to 42°C, the thermolabile repressor undergoes a change which results in its inactivation. Since, pCE30 also expresses the *cI857* gene, in transformed cells growing at 30°C, the P_R P_L promoters are repressed. A shift in temperature to 42°C results in initiation of transcription from the tandem promoters and subsequent gene expression. A similar strategy is used for control of expression in other systems, including the pEX vectors discussed in Chapter 5. A further consideration is the state of the protein produced. The stability and solubility of expressed polypeptides, especially those alien to the host cell, or different from their native form, or simply produced in abnormally large quantities, are discussed in Chapter 3.

2.2.1 Cloning Strategies

The cloning strategy used to generate the homeobox expression vector is outlined below and summarised in Figure 2.3.

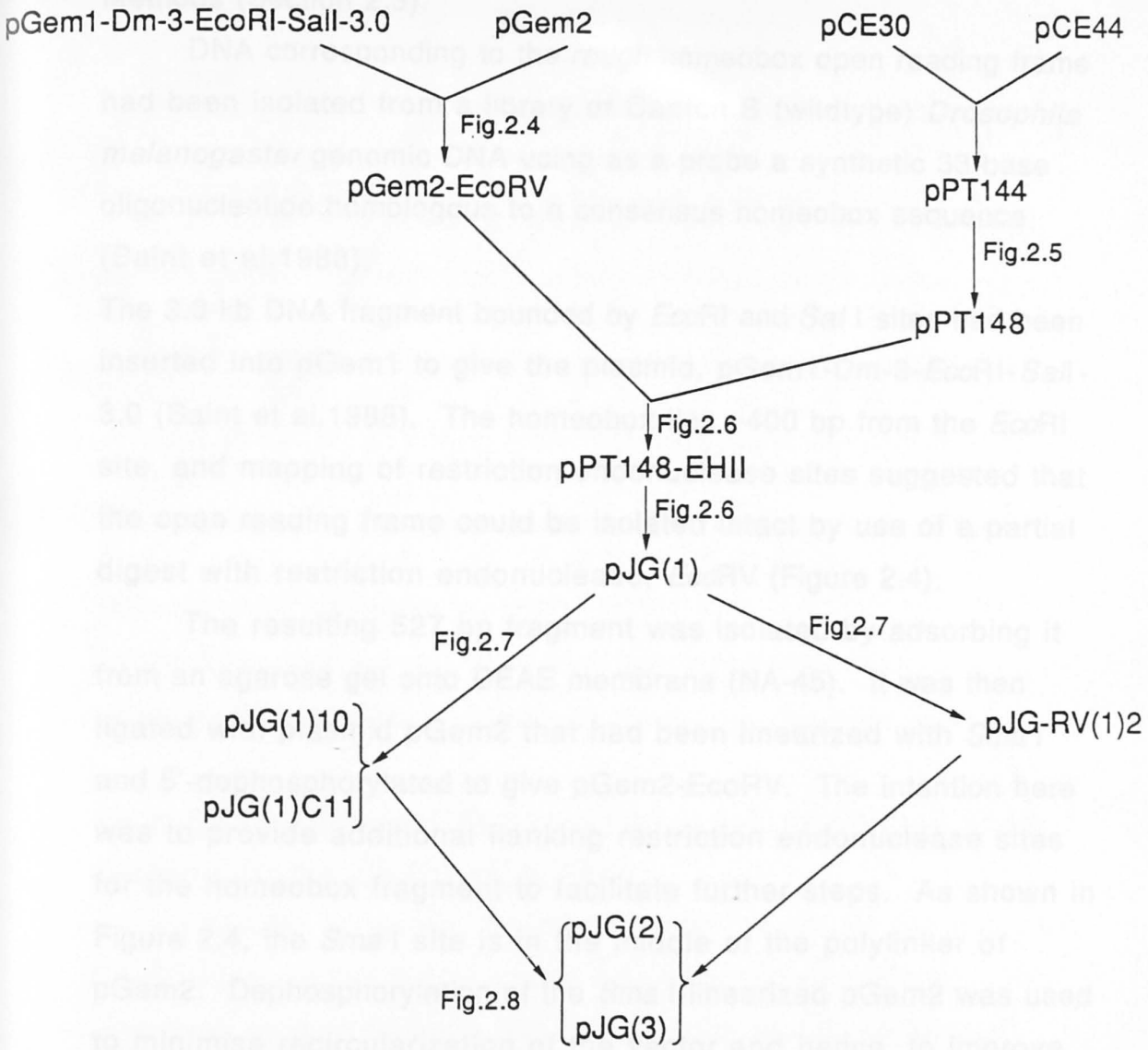


Figure 2.3 : Summary of plasmid constructions in Chapter 2.

Details of experimental methods are given under Materials and Methods (Section 2.3).

DNA corresponding to the *rough* homeobox open reading frame had been isolated from a library of Canton S (wildtype) *Drosophila melanogaster* genomic DNA using as a probe a synthetic 33-base oligonucleotide homologous to a consensus homeobox sequence (Saint et al.1988).

The 3.0-kb DNA fragment bounded by *Eco*RI and *Sal* I sites had been inserted into pGem1 to give the plasmid, pGem1-Dm-3-*Eco*RI-*Sal*I-3.0 (Saint et al.1988). The homeobox lies ~400 bp from the *Eco*RI site, and mapping of restriction endonuclease sites suggested that the open reading frame could be isolated intact by use of a partial digest with restriction endonuclease, *Eco*RV (Figure 2.4).

The resulting 527 bp fragment was isolated by adsorbing it from an agarose gel onto DEAE membrane (NA-45). It was then ligated with plasmid pGem2 that had been linearized with *Sma* I and 5'-dephosphorylated to give pGem2-EcoRV. The intention here was to provide additional flanking restriction endonuclease sites for the homeobox fragment to facilitate further steps. As shown in Figure 2.4, the *Sma* I site is in the middle of the polylinker of pGem2. Dephosphorylation of the *Sma* I-linearized pGem2 was used to minimise recircularization of the vector and hence, to improve the efficiency of the desired ligation. The ligation mixture was then used to transform competent *E. coli* cells (strain RR1). Since all of the plasmids used in this work bear the β-lactamase (Amp^R) gene, selection of plasmid-bearing transformants was by their resistance to the antibiotic, ampicillin (at 50μg/ml). Small-scale plasmid preparations of transformants were examined. Digestion with *Eco*RI showed the presence of the homebox fragment, and its orientation in the polylinker was determined following digestion with *Bst* N1 (Figure 2.4).

```
      EcoRV
TGTAAGGATATCAATGTATTTGACTTTCAGTGGCTCGCAGGCGACGCAAGGAGGGAAGAC 383
                     *  L  S  V  A  R  R  R  R  K  E  G  R   12
            AvaII
AACGCCGCCAGAGGACCACTTTCAGCACAGAGCAGACGCTTCGCCTGGAGGTGGAGTTCC 443
Q  R  R  Q  R  T  T  F  S  T  E  Q  T  L  R  L  E  V  E  F   32

ATCGGAACGAGTACATCTCCAGGAGTCGTCGCTTCGAGCTGGCCGAAACGCTGCGTCTGA 503
H  R  N  E  Y  I  S  R  S  R  R  F  E  L  A  E  T  L  R  L   52
               BglII
CCGAAACGCAAATCAAGATCTGGTTCCAGAATCGCAGGGCCAAGGACAAACGCATTGAAA 563
T  E  T  Q  I  K  I  W  E  Q  N  R  R  A  K  D  K  R  I  E   72

AGGCTCAGATCGATCAGCACTACAGGTGGGTTTCCCCTTTAGCTAATTTGAATCGTTTTT 623
K  A  Q  I  D  Q  H  Y  R  W  V  S  P  L  A  N  L  N  R  F   92
                  Eco RV
TCTGTTTCATATTTACAAAGATATCTTCGATTTTGATTGAGTTAAATAACATTAAATGAA 683
P  C  F  I  F  T  K  I  S  S  I  L  I  E  L  N  N  I  K  *  111

CATAAATTAAAGGTTGATGAATAGCTCGTGCAGAGGATTTTAACTGGTTATTTATACCTG 743

CATTTTCGAAACGGAAGTAATTTAGTGAAATGTCACTAGTTTAATTGATGGCCAACATAA 803
                                                    Eco RV
AATTAACACGAAATTAATTTTACGACATTTATGCGCACACAGACCCAATTAGACGATATC 863
```

Figure 2.4a): Nucleotide sequence of the homeobox open reading frame (Saint et al.1988).

The DNA sequence is numbered from the *Eco* RI site (which is 3028 in the complete *rough* gene sequence shown in Figure 1.1b). The open reading frame is 333 bp in length, and its conceptual translation is shown in the one-letter amino acid code. The residues comprising the homeodomain are underlined. Restriction enzyme sites used in the manipulation of this fragment in this chapter are indicated. The exon-intron splice sites at 352 and 589 are underlined.

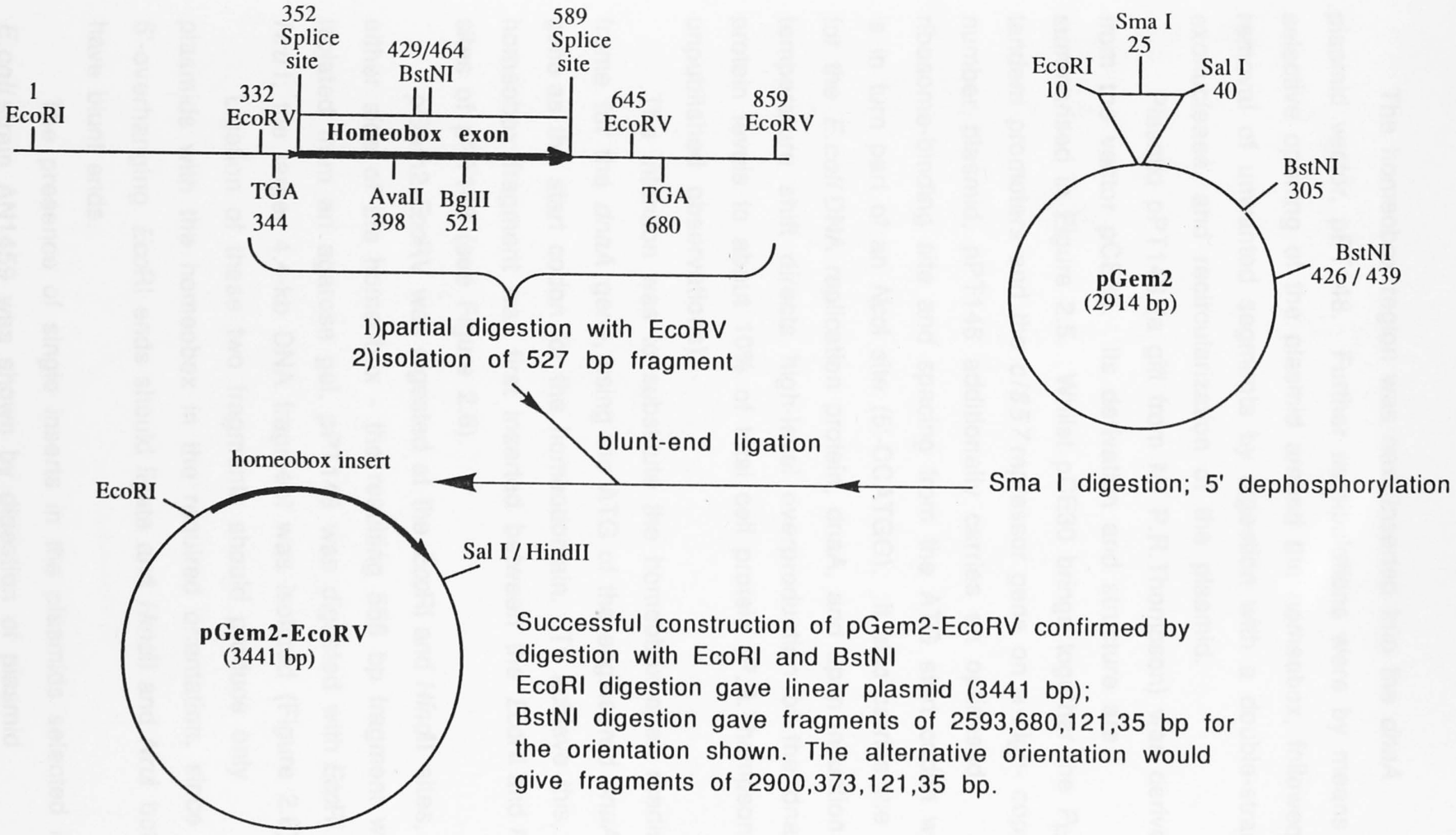


Figure 2.4b): Restriction Map of Homeobox region and its insertion into pGem2 to give pGem2-EcoRV.

The homeobox region was next inserted into the *dnaA* plasmid vector, pPT148. Further manipulations were by means of selective opening of the plasmid around the homeobox, followed by removal of unwanted segments by digestion with a double-stranded exonuclease, and recircularization of the plasmid.

Plasmid pPT148 (a gift from Mr P.R.Thompson) was derived from the vector pCE30. Its derivation and structure are summarised in Figure 2.5. Whilst pCE30 brings together the $P_R P_L$ tandem promoters and the *cI857* repressor gene on a high- copy-number plasmid, pPT148 additionally carries an optimised ribosome-binding site and spacing from the ATG start codon which is in turn part of an *Nco*I site (5'-CCATGG). It also carries the gene for the *E.coli* DNA replication protein, dnaA, and upon induction by temperature shift directs high-level overproduction of the dnaA protein levels to about 10% of total cell protein (P.R.Thompson, unpublished observations).

The intention was to substitute the homeobox open reading frame for the *dnaA* gene, using the ATG of the engineered *dnaA* gene as the start codon for the homeodomain. To achieve this, the homeobox fragment was first inserted between the *Eco*RI and *Nru*I sites of pPT148 (see Figure 2.6).

pGem2-*Eco*RV was digested at the *Eco*RI and *Hind*II sites, on either side of the homeobox - the resulting 558 bp fragment was isolated from an agarose gel. pPT148 was digested with *Eco*RI and *Nru* I; the larger 4.4-kb DNA fragment was isolated (Figure 2.6).

Ligation of these two fragments should produce only plasmids with the homeobox in the required orientation, since the 5'-overhanging *Eco*RI ends should ligate and *Hind*II and *Nru*I both have blunt ends.

The presence of single inserts in the plasmids selected in *E.coli* strain AN1459 was shown by digestion of plasmid

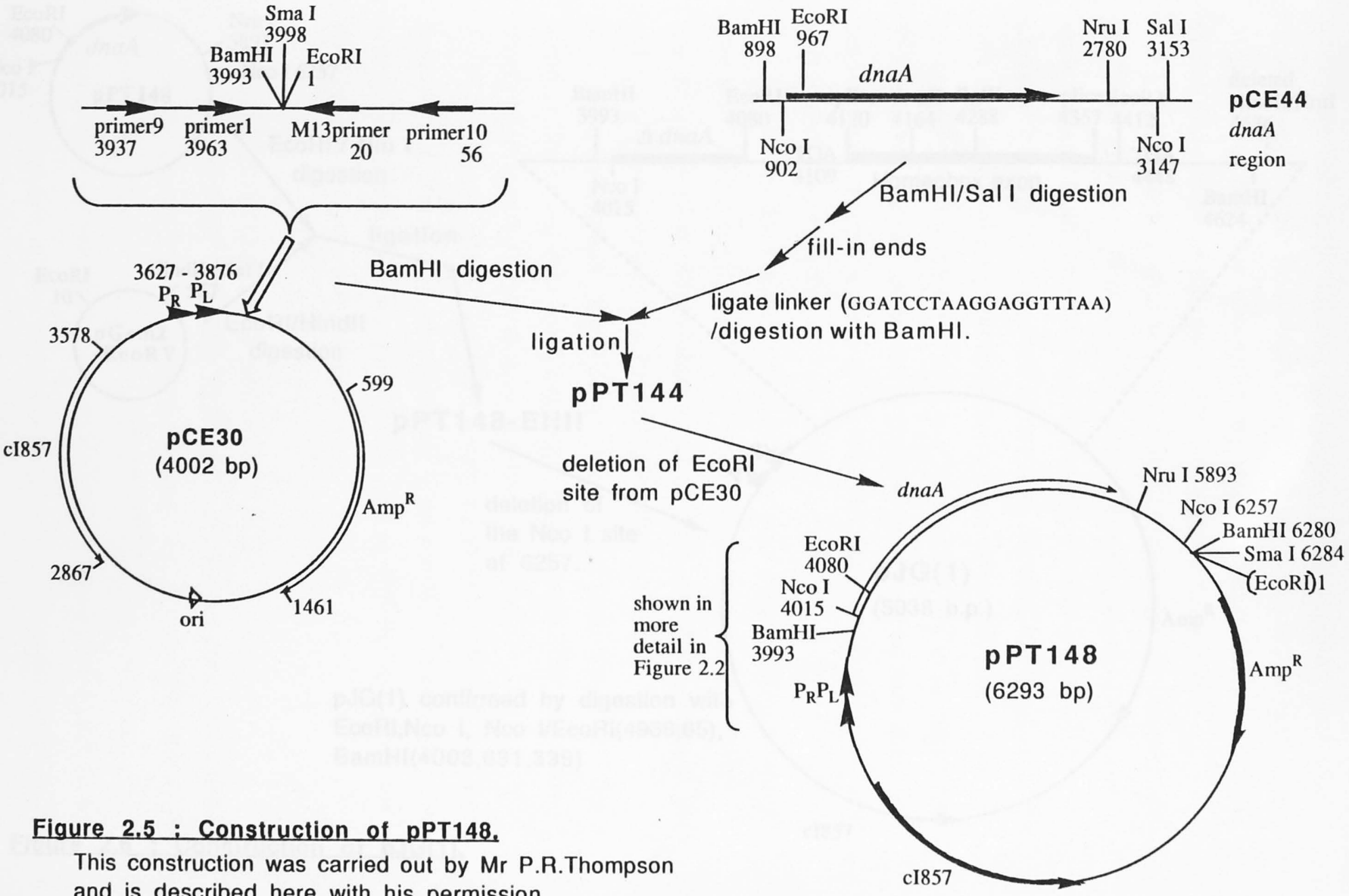


Figure 2.5 : Construction of pPT148.

This construction was carried out by Mr P.R.Thompson and is described here with his permission.

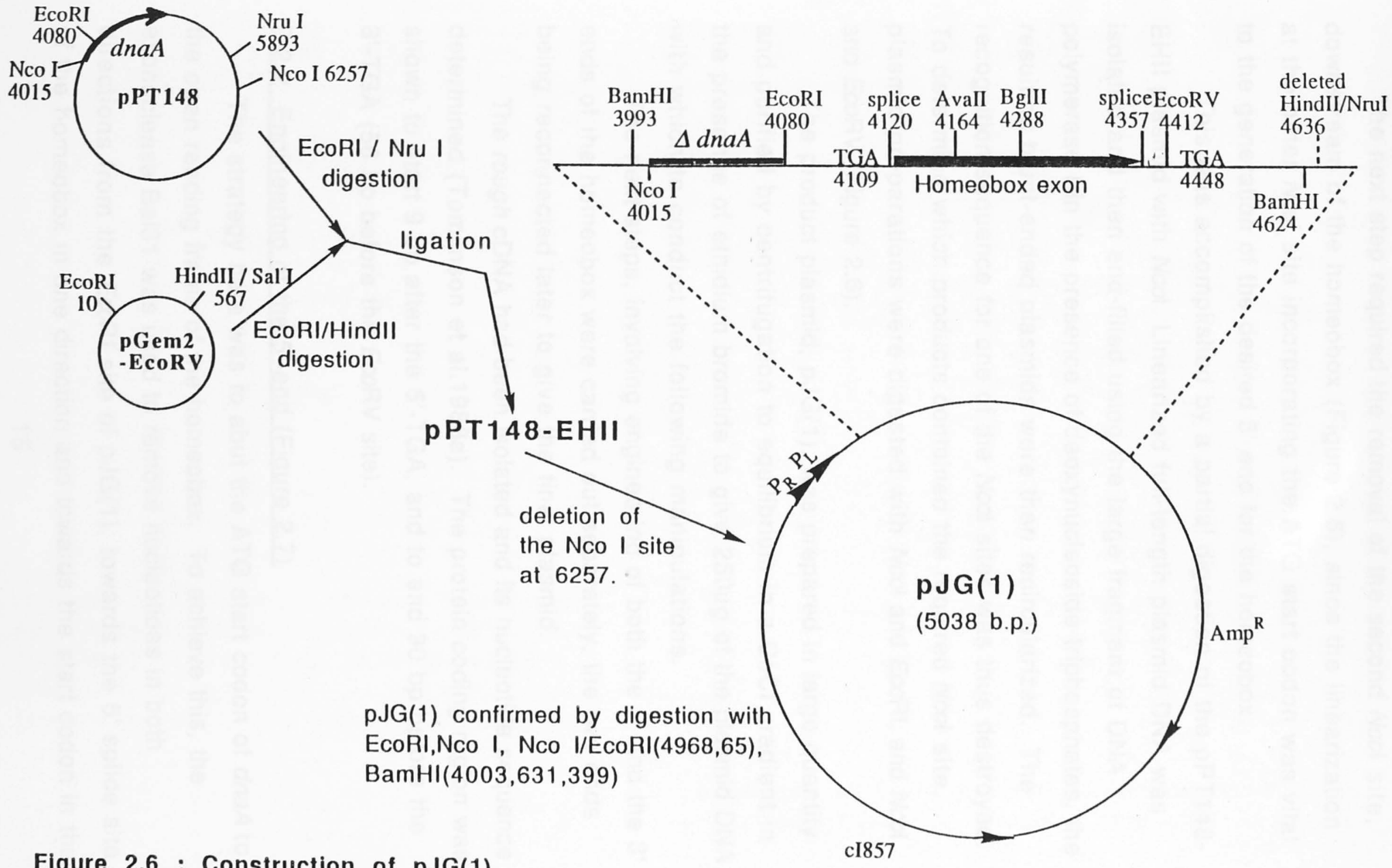


Figure 2.6 : Construction of pJG(1).

preparations with *Bam*HI and *Eco*RI.

The next step required the removal of the second *Nco*I site, downstream of the homeobox (Figure 2.6), since the linearization at the other *Nco*I site incorporating the ATG start codon was vital to the generation of the desired 5' end for the homeobox.

This was accomplished by a partial digestion of the pPT148-EHII plasmid with *Nco*I. Linearized full-length plasmid DNA was isolated and then end-filled using the large fragment of DNA polymerase I in the presence of deoxynucleoside triphosphates, the resulting blunt-ended plasmids were then recircularized. The recognition sequence for one of the *Nco*I sites was thus destroyed. To determine which products contained the required *Nco*I site, plasmid preparations were digested with *Nco*I and *Eco*RI, and *Nco*I and *Eco*RV (Figure 2.6).

The product plasmid, pJG(1), was prepared in large quantity and purified by centrifugation to equilibrium in a CsCl gradient in the presence of ethidium bromide to give 250μg of the plasmid DNA with which to conduct the following manipulations.

The next steps, involving engineering of both the 5' and the 3' ends of the homeobox were carried out separately, the two ends being reconnected later to give the final plasmid.

The *rough* cDNA had been isolated and its nucleotide sequence determined (Tomlinson et al.1988a). The protein coding region was shown to start 9 bp after the 5'-TGA, and to end 90 bp before the 3'-TGA (56 bp before the *Eco*RV site).

2.2.2 Engineering of the 5'-end (Figure 2.7)

The strategy here was to abut the ATG start codon of *dnaA* to the open reading frame of the homeobox. To achieve this, the exonuclease Bal31 was used to remove nucleotides in both directions from the *Eco*RI site of pJG(1), towards the 5' splice site of the homeobox in one direction and towards the start codon in the

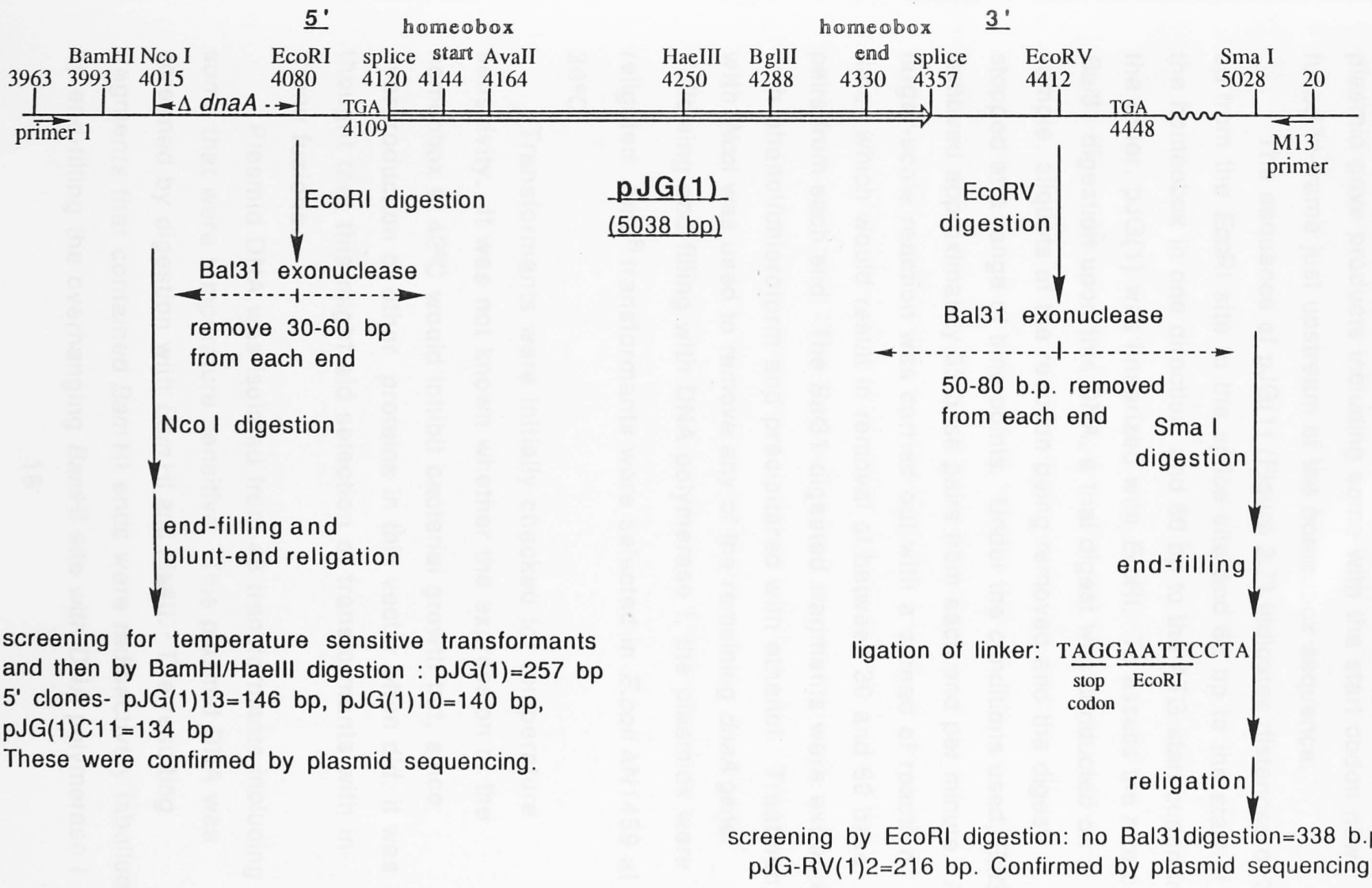


Figure 2.7 : The engineering of the 5' and 3' ends of the *rough* Homeobox exon in pJG(1).

other. Subsequent digestion with *Nco*I removed any remaining *dnaA* sequence and, following end-filling, recircularization of the plasmid gave products including some with the start codon now fused in-frame just upstream of the homeobox sequence.

The sequence of pJG(1) (Figure 2.7) indicates distances of 38 bp from the *Eco*RI site to the splice site and 62 bp to the start of the homeobox in one direction, and 60 bp to the ATG start codon in the other. pJG(1) was linearized with *Eco*RI. To assess the rate of *Bal*31 digestion upon this DNA, a trial digest was conducted on a sample, aliquots of the reaction being removed and the digest stopped at a range of time points. Under the conditions used, *Bal*31 removed approximately 35 base pairs from each end per minute. A larger-scale reaction was carried out with a spread of reaction times which would result in removal of between 30 and 60 base pairs from each end. The *Bal*31-digested fragments were extracted with phenol/chloroform and precipitared with ethanol. Treatment with *Nco*I was used to remove any of the remaining *dnaA* gene. Following end-filling with DNA polymerase I, the plasmids were religated. Amp^R transformants were selected in *E.coli* AN1459 at 30°C.

Transformants were initially checked for temperature sensitivity. It was not known whether the expression of the homeobox at 42°C would inhibit bacterial growth but, since overproduction of other proteins in this vector often did, it was thought that this might aid selection of transformants with in-frame fusions.

Plasmid DNA was isolated from 48 transformants including some that were temperature sensitive. This plasmid DNA was screened by digestion with *Bam* HI and *Hae*III. The resulting fragments that contained *Bam* HI ends were radioactively labelled by end-filling the overhanging *Bam*HI site with DNA polymerase I

and [α-^{32}P]dATP and were separated by electrophoresis on thin acrylamide gels. Sizes were determined by reference to similarly labelled size standards (*Sau*3A digest of pCE30) (Figure 2.7). Several plasmids gave labelled fragments between 130 and 150 bp. Their nucleotide sequences were determined by the method of Sanger et al.(1982) using NaOH-denatured plasmid DNA as template, primer 1 of Elvin et al.(1990), and [α-^{32}P]dATP. Denaturation of the double-stranded plasmid gave single-stranded DNA to which a primer oligonucleotide could be annealed (the primer site shown in Figure 2.7 is upstream of the RBS, and directs synthesis over the start codon and succeeding sequence). The dideoxy chain termination sequencing reaction then proceeded with the large fragment of DNA polymerase I synthesising a new DNA strand from the primer site, incorporating [α-^{32}P]dATP. Reactions were terminated by substitution of one of the four dNTP's with a ddNTP. The reaction mixture was denatured and electrophoresed through an 8M urea/5% polyacrylamide gel. The dried gel was autoradiographed and gave the sequence of each plasmid in the region of the start codon.

It emerged from this screening that only transformants containing plasmids with in-frame fusions were temperature sensitive for growth at 42^{o}C. Three plasmids which met the required aims, namely pJG(1)13, pJG(1)10 and pJG(1)C11, were fully characterized. All three had the ATG start codon fused in-frame to the coding region between the splice site and the start of the homeobox. These sequences are given in Figure 2.8. Note that in pJG(1)C11 the *Nco*I site at the start codon was regenerated, which may be of use in future sub-cloning into other vectors.

2.2.3 Engineering of the 3' End (Figure 2.7).

In an analogous manner, *Bal*31 was used to remove

nucleotides between the *Eco*RV site and the 3' splice site for the exon. In addition, a synthetic oligonucleotide carrying a termination codon was introduced to terminate translation of the sequence to be expressed.

Figure 2.7 indicates 56 bp of DNA between the *Eco*RV site and the splice site, and 83 bp to the end of the homeobox. Plasmid pJG(1) was digested with *Eco*RV. A trial *Bal*31 digest gave a rate of 23 base pairs per end per minute. The scaled-up digest was designed to remove from 50 to 80 bp from each end.

The *Bal*31-digested fragments were then digested with the restriction endonuclease, *Sma*I, to remove DNA foreign to the vector pCE30, and to facilitate sequence determination using the M13 universal primer as described (Elvin et al.1990) (Figure 2.7). The fragments were now end-filled to ensure the remaining *Bal*31-digested end was blunt. Before reclosing the plasmid, a 12-mer oligonucleotide bearing the TAG stop codon and *Eco*RI site was introduced. This ligation entailed mixing the plasmid with around 100-fold molar excess oligonucleotide linker in the presence of ligase. Use of unphosphorylated linker ensured insertion of a single copy of it between the blunt ends. Heating of the mixture to 65°C to remove unligated annealed linker followed by slow cooling permitted renaturation of the overhanging 12-mer strands of the linker and recircularization of the plasmid.

This mixture was then used to transform AN1459, with selection for Amp^R. Transformants were screened for an in-frame termination codon in the desired region. This was carried out by digestion of plasmid isolated from 12 transformants, followed by end-labelling of the fragments. These were electrophoresed, with standards, through a 5% polyacrylamide gel. The autoradiogram was scrutinised for appropriately-sized fragments (between 204 and 231 bp, Figure 2.7). Nucleotide sequencing reactions were

carried out as above, using the M13 universal primer with termination only by ddATP. The A-track was then matched against the known sequence; this allows more plasmids to be screened more rapidly than does full sequencing of each. From a combination of these two methods, two promising plasmids were identified, and their sequences were determined as above. One of them, pJG-RV(1)2, was found to have an in-frame termination codon 12 bp downstream of the end of the homeobox and 15 bp upstream of the splice site.

2.2.4 Uniting the 5'- and 3'- Ends. (Figure 2.8).

With plasmids bearing the appropriately engineered 5' and 3' ends, it was necessary to bring them together to give the final product.

Initially, it had been intended to digest at the *Bgl*II site in the middle of the homeobox together with the one in the vector to produce two easily-ligatable fragments. However, sequencing demonstrated that pJG-RV(1)2 contained a new *Bgl*II site, created at the juncture of the 3'-end with the stop codon. Thus, the entire 3' end was lost as a 55 bp fragment by *Bgl*II digestion. Since the small size of this fragment made it difficult to monitor a partial digestion, *Ava* II digestion was chosen as the next best alternative. There was one *Ava*II site in the homeobox, and two well-spaced sites in the vector region (Figure 2.8).

The 1.4 kb fragment from a partial *Ava*II-digest of pJG-RV(1)2 was purified from an agarose gel and ligated with the 2.8 kb fragment from the *Ava*II digestion of the 5'-ATG-engineered plasmids. Plasmids in transformants were checked for 232 bp fragments following digestion with *Bam*HI and *Eco*RI and end-labelling. Complete nucleotide sequence of the homeobox insert in 6 of these was determined (see photograph of pJG(2) sequence in

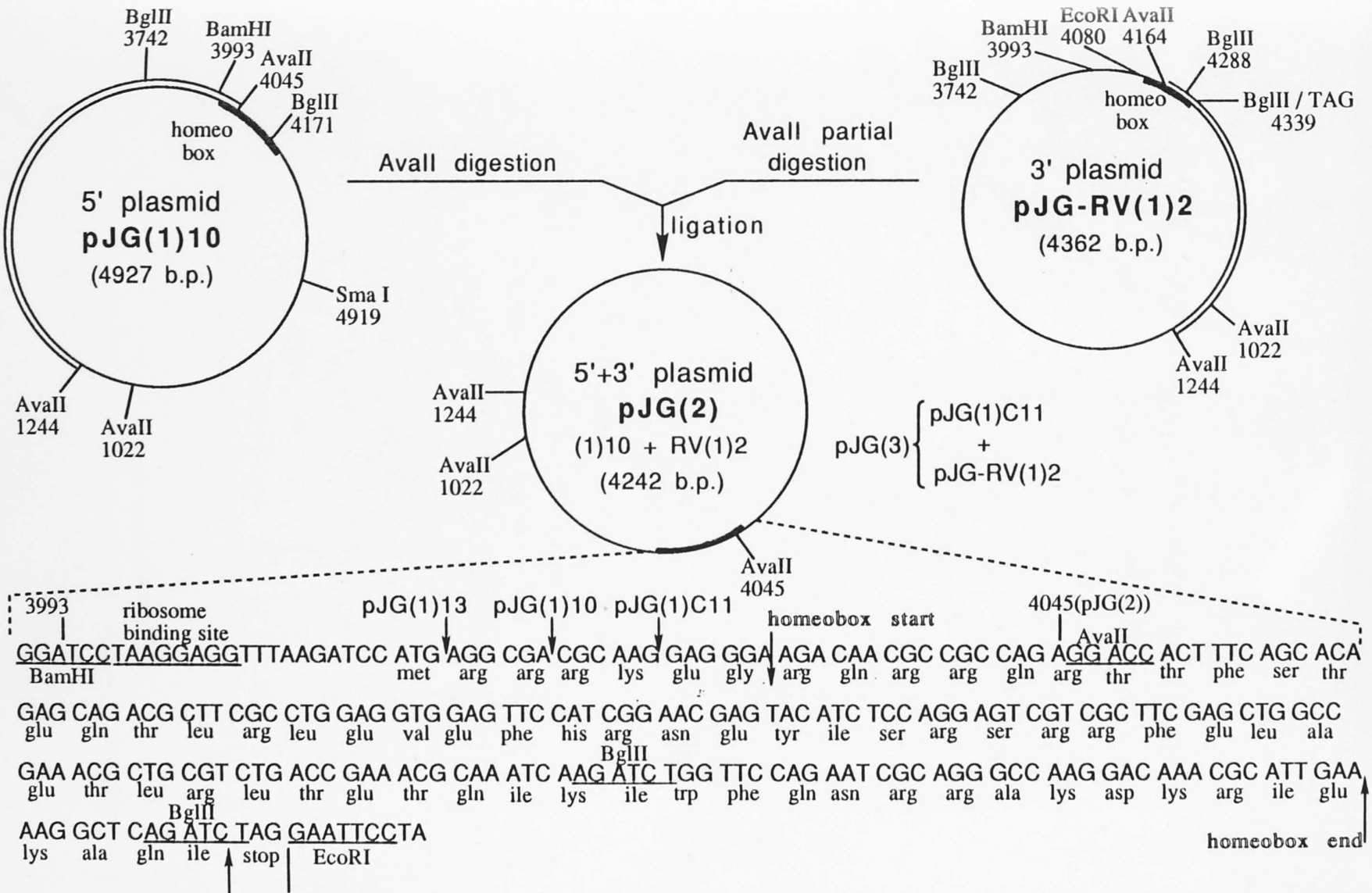


Figure 2.8 : Uniting the 5' and 3' ends. Sequence of the final constructs.

Figure 2.9: Autoradiogram showing the nucleotide sequence of the *rough* homeobox region within pJG(2).

The plasmid was sequenced using the method described in Section 2.3.12. It should be read, 5' to 3', from bottom to top, left to right. The DNA sequence and the homeodomain construct it is predicted to encode are given in Figures 2.8 and 3.1.

The highlighted regions are as follows (the numbering indicates the distance from the deleted *Eco* RI site of pPT148 (see Figure 2.5)):

1: TAAGGAGG - the ribosome binding site (3997).

2: ATG - the start codon of the pJG(2)*rough* homeodomain protein (4015).

3: AGA - the first codon of the homeodomain (4030).

4: GGACC - *Ava*II site (4046).

5: AGATCT - *Bgl*II site (4169).

6: GAA - the last codon of the homeodomain (4207).

7: AGATCT - *Bgl*II site (4216).

8: TAG - the stop codon of the pJG(2) *rough* homeodomain protein (4221).

9: GAATTC - *Eco* RI site (4224).

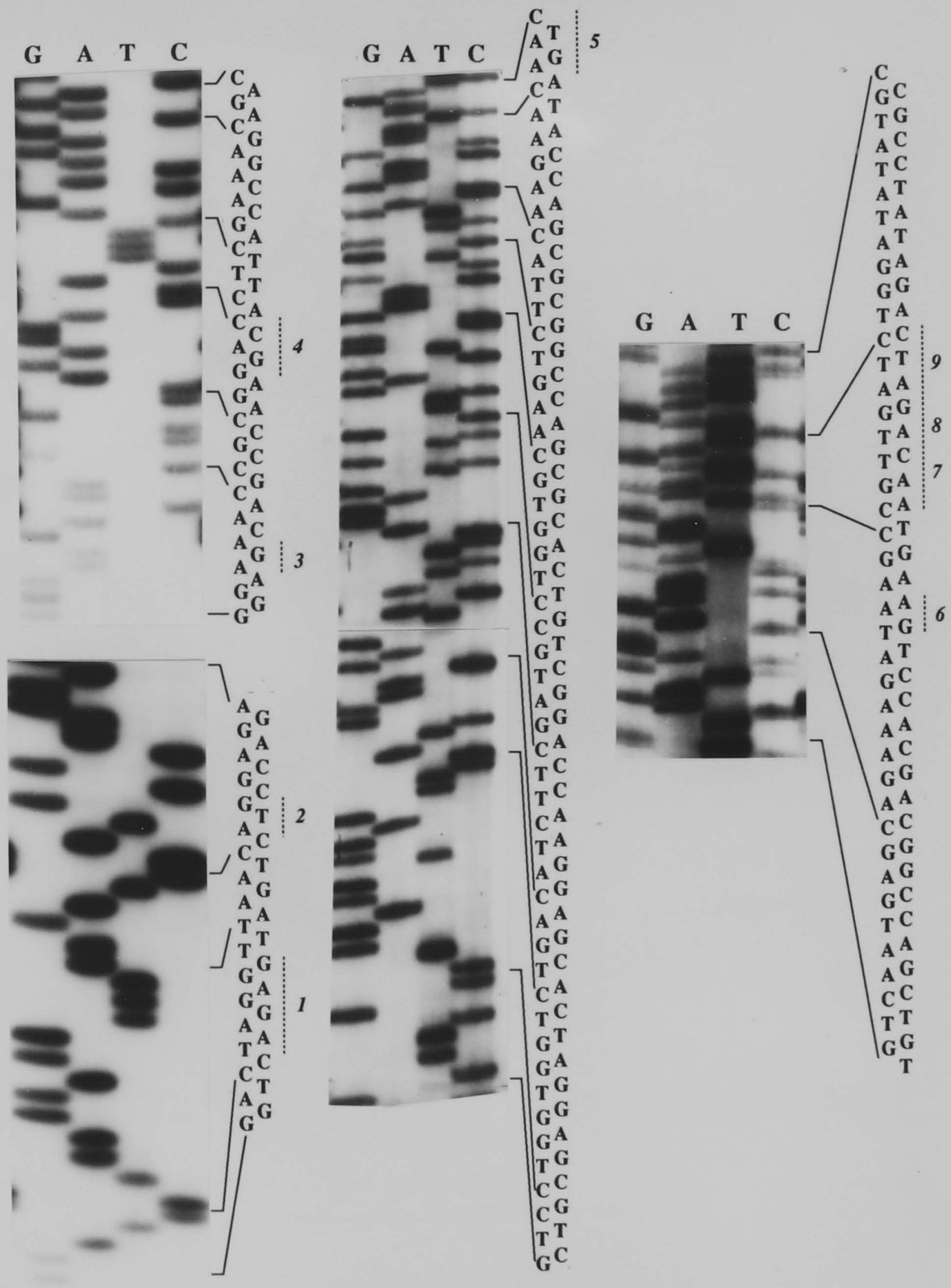


Figure 2.9: Autoradiogram showing the nucleotide sequence of pJG(2) using the dideoxy method.

Figure 2.9). Two plasmids were thus obtained, pJG(2), representing the joining of pJG-(1)10(4) and pJG-RV(I)2, and pJG(3), in which pJG-(1)C11 was united to pJG-RV(1)2.

Expression of the homeodomain polypeptide directed by these plasmids is described in Chapter 3.

2.3 Materials and Methods

Most of the following methods are fairly standard and may be found in such books of Molecular Biology techniques as Maniatis et al.(1982), Berger and Kimmel (1987), and Dillon et al.(1985).

2.3.1 Restriction Endonuclease Digests

DNA was mixed with a restriction endonuclease in an appropriate buffer, and incubated at the relevant temperature. Most of the restriction endonucleases used in this work were supplied by Boehringer Mannheim Biochemicals and more complete details of their activity, digestion buffers and conditions are given in the relevant sales catalogue.

A double digest simply entailed mixing the DNA with two restriction endonucleases in a buffer which permitted both enzymes to operate effectively (e.g. a 50:50 mixture of their individual buffers).

A partial digest required the use of conditions which resulted in an incomplete digest of the available enzyme sites in the DNA population, e.g. less than the optimal amount of enzyme and a shorter incubation time. The reaction was stopped by placing the reaction vessel on dry ice.

2.2.2 Agarose Gel Analysis of DNA

A hot solution of agarose (usually in the range 0.7 to 1.5%) was made up in running buffer (TBE = 90mM Tris, 90mM borate, 2.5mM

EDTA) and then poured into a perspex mould to set. A comb was placed at the cathode end to introduce a given number of slots in the gel for loading DNA samples, and the solution left to set.

The mould doubled as an electrophoresis chamber; enough running buffer was added to cover the gel and the electrodes, the comb was then removed and DNA samples were loaded into the slots. DNA samples usually had at least a 1/10th volume of a bromophenol blue-glycerol solution (0.2% bromophenol blue (w/v), 50% glycerol (v/v)) added to them prior to loading to act as a dye marker, to assess the course of electrophoresis. Electrophoresis was commenced by applying a voltage (up to around 100 volts) across the electrodes. The negatively-charged DNA particles migrated towards the anode.

A molecular weight standard was usually included in one slot. This was λ DNA digested with the restriction enzyme, *Hind*III, to generate the fragments 23,130 / 9,416 / 6,682 / 4,361 / 2,322 / 2,027 / 565 and 125 base pairs.

DNA fragments were visualised using a UV-fluorescing dye, ethidium bromide (EtBr), which binds to the DNA molecule. EtBr (at 0.5 μg/ml) was usually added to the gel solution and the running buffer. The DNA-EtBr complexes were visualised by using an ultraviolet source, usually by laying the gel on a transilluminator, which provides an even U.V. source from below, and the gel was photographed with a polaroid camera bearing a red filter.

Materials

Agarose (BioRad); T.B.E. running buffer: 10 x's solution=108 g Tris, 56g Boric Acid, 9.4g EDTA per litre; Polaroid MP4 Land Camera; Polaroid film Type 57-3000 ASA; Red filter: Kodak Wratten Gelatin filter.

2.3.3 Isolation of DNA bands from Agarose Gels using NA-45 Paper (DEAE Membrane).

The DNA sample was electrophoresed through an agarose gel. DNA bands were visualised and those of interest identified. An appropriately sized piece of NA-45 paper (Schliecher and Schuell) was soaked in buffer and placed against the gel slice containing the DNA band. When the current was re-applied, the DNA migrated onto the paper to which it adhered. DNA was eluted off the NA-45 paper by immersing it in elution buffer (1M NaCl, 50mM Arginine) and heating it at 65°C for >1 hour. The paper was then removed, and the DNA precipitated from the solution by adding two volumes of ethanol and freezing in dry ice. The tube was then spun at high speed for 10 minutes at 4°C and the resulting supernatant removed. The pellet was rinsed with 70% ethanol, dried under vacuum and resuspended in a small quantity of TE (10mM Tris (pH usually 7.5),1mM EDTA).

2.3.4 Phosphatasing DNA.

The terminal phosphate group of plasmid DNA was removed to prevent its recircularisation during ligation reactions, involving the addition of non-phosphorylated DNA fragments. A typical reaction was :-10µl *Sma*I-cut pGem2 + 10µl 10 x CIP buffer + 79µl H_2O +1µl calf intestinal phosphatase. Incubated at 37°C for 30 minutes, then at 55°C for 15 minutes. 12µl 10 x STE and 5µl 10% SDS was added and the mixture incubated at 70°C for 10 minutes to inactivate the phosphatase. The solution was deproteinized by Phenol/Chloroform extraction. This entailed mixing it with 1 volume of phenol and leaving it for 5 minutes before centrifugation. The upper aqueous layer was then removed and mixed with 1 volume of phenol/chloroform and centrifuged. Finally, the aqueous layer was mixed with 1 volume of chloroform

and centrifuged.

The DNA was then precipitated from the aqueous layer by addition of a 1/10th volume of 3M sodium acetate, 2 volumes of ethanol, and by freezing the solution at -20°C for at least half an hour. The eppendorf tube was then centrifuged for 10 minutes, at 4°C, and the pellet washed carefully with 70% ethanol, dried *in vacuo* and resuspended in 20µl TE.

Materials

Calf intestinal alkaline phosphatase (Pharmacia); 10 x CIP buffer: 0.5M Tris pH 9.0, 10mM $MgCl_2$, 1mM $ZnCl_2$, 10mM spermidine; 10 x STE buffer: 100mM Tris pH8.0, 1M NaCl, 10mM EDTA; 10% SDS = 10% sodium dodecyl sulphate; Phenol solution: crystalline phenol and equilibrated with TE until saturated, 0.1% 8-hydroxyquinoline is added to prevent oxidation; Phenol/chloroform solution: phenol solution mixed with chloroform and isoamyl alcohol in the volume ratio 24 : 24 : 1.

2.3.5 Ligation.

Here, the matching ends (blunt, or homolgous overhangs from a restriction endonuclease digestion) of DNA fragments were covalently joined by the ligase enzyme. A typical blunt-end ligation: 5µl *Eco*RV-cut insert (500 bp, 0.2µg) + 1µl pGem2, *Sma*I-cut and phosphatased (3 kb,1µg) + 1µl 10 x blunt ligation buffer + 1µl T4 DNA ligase + 2µl water. Incubated at 14°C overnight. Homologous ends: 2µl *Eco*RI /*Nru*I digested pPT148 (6 kb, 0.3µg) + 2µl *Eco*RI /*Hind*II digested insert (550 bp,0.3µg) + 1µl 10 x T4 buffer + 1µl T4 DNA ligase + 4µl water. Incubate at 30°C for 2 hours.

Materials

T4 DNA ligase (1 unit/µl) BRESA; 10 x blunt ligation buffer: 250mM Tris pH 7.5, 50mM $MgCl_2$, 25% (w/v) Polyethylene glycol 8000, 5mM

DTT, 4mM ATP; 10 x T4 ligation buffer: 250mM Tris pH 7.5,100mM $MgCl_2$, 100mM DTT, 4mM ATP.

2.3.6 Transformation of *E.coli* cells.

Plasmid DNA was introduced into *E.coli* cells by making the cells "competent" before adding the DNA. Preparation of competent cells: 1. 5ml LB culture of AN1459 was grown overnight at 37°C; 2. this was made up to 50ml and grown for a further 1-2 hours at 37°C; 3. the exponential phase cells were chilled for 20 minutes at 4°C; 4. they were centrifuged 6000g at 4°C; 5. the cell pellet was resuspended in 10ml 0.1M $MgCl_2$ to wash them, and then re-centrifuged; 6. the pellet was resuspended in 10ml 0.1M $CaCl_2$ and left for >1hour at 4°C; 7. this was re-centrifuged, and the pellet resuspended in 5ml 0.1M $CaCl_2$ /15% glycerol. Aliquots could be used immediately for transformation or stored indefinitely at -70°C and thawed at 4°C as needed.

Transformation of competent cells: 1. 200µl aliquot of competent AN1459 cells was mixed with the relevant DNA solution at 4°C; 2. after 20 minutes, the aliquots were heat-shocked in a waterbath at 30 or 37°C for 2 minutes; 3. 1ml of LB was added and the cells were left for a further 15 minutes; 4. the cells were spun down and resuspended in a small volume of LB. This was then placed as a drop on an LB Amp plate and spread using a bent glass rod.; 5. the plate was incubated overnight at 30 or 37°C.

Materials

E.coli strains: AN1459 = *ilvC thr leu srl*::Tn*10 recA* (lab stock); RR1 = F^- *hsdS20 proA lacY galK rpsSL20 xyl15 mtl1 supE44* (Maniatis et al.1982); Luria-Bertani broth (LB): 10g tryptone, 5g yeast extract, 5g NaCl, 2.5ml NaOH (1M) per litre. For agar plates, 15g agar was added per litre LB. 2ml Ampicillin solution (25mg/ml) was added per litre for LB-Amp.

2.3.7 Screening of Transformants

Transformants with plasmids bearing the β-lactamase (Amp^R) gene were screened by growth on nutrient agar plates with ampicillin. Temperature-inducible plasmid expression vectors were screened for poor growth at 42°C as an indication of plasmids expressing a cloned gene. Final screening entailed selecting a dozen or so clones and making preparations of their plasmid DNA.

2.3.8 Plasmid Preparations.

A number of methods were used for the preparation of plasmid DNA according to the quantity and purity required.

Size Check: For a quick check of changes in the plasmid size due to the addition of an insert, a small sample of the bacterial clone was lysed with SDS and loaded fairly directly onto an agarose gel - this gives a quick assessment of the size of the plasmid relative to a control, i.e. bigger/smaller/same: Transformants were toothpicked to another petri dish, and the remainder removed to an eppendorf tube with 25µl of cracking buffer (50mM NaOH; 0.5% SDS; 5mM EDTA) and heated at 70°C for 30 minutes; 3µl of 25% sucrose; 50mM Tris (pH8); 50mMEDTA, and 3µl bromophenol blue/glycerol loading buffer were then added; The extract was electrophoresed on an agarose gel slots, with undigested plasmid DNA controls.

Rapid lysate / Miniprep: This method usually gave a few µg of DNA which could be digested to look for fragments: 1. clones were streaked out on 6-sector plates, and grown up overnight at 30 or 37°C. The grown cells were scraped off the agar surface and resuspended in 200µl of lysozyme solution (50mM glucose;10mM EDTA; 25mM Tris pH 8.0; 1mg/ml lysozyme). Incubated for 5 minutes on ice then 0.4ml of NaOH-SDS solution was added (0.2M NaOH; 1% SDS) and tubes were mixed by inversion. Incubated for 5 minutes on ice then 0.3ml KOAc solution (3M potassium acetate;

28.5% (v/v) glacial acetic acid) added. The mixture was frozen on dry ice, thawed at room temperature and centrifuged for 10 minutes at top speed in the bench microfuge. 0.75ml of the supernatant was removed and 0.45ml isopropanol mixed with it at room temperature for 20 minutes to precipitate DNA. This was centrifuged for 10 minutes and the pellet was washed carefully with 1ml 70% ethanol, dried by vacuum and resuspended in 50µl TE + 20 µg/µl RNase A. This was incubated for 15 minutes at 37°C and 15 minutes at 70°C. 5µl of this solution was usually enough for one lane on an agarose gel. However, it was often necessary to phenol/chloroform extract the solution in order to obtain DNA clean enough for restriction enzyme digests.

Midi-prep: This method gave around 100µg of plasmid DNA, pure enough for enzyme digests and plasmid sequencing: A 40ml overnight culture of the clone was grown up, and the cells were centrifuged at 6000g for 5 minutes at 4°C, and resuspended in 1ml of 25mM Tris pH8.0, 10mM EDTA. 2ml of 0.2M NaOH/1% SDS solution was added, mixed and left on ice for 5 minutes. 1.5ml of the KOAc solution (3M KOAc; 28.5% (v/v) glacial acetic acid) was then added and left for 10 minutes on ice. This was centrifuged at 15,000g for 15 minutes at 4°C, and the supernatant was removed carefully and mixed with 50µl of a 10mg/ml RNase A solution and 20 µl 10mg/ml Proteinase K at 37°C for 1 hour. 2 volumes of ethanol were added, and the solution frozen at -20°C, then centrifuged at 15,000g for 15 minutes at 4°C. The DNA pellet was resuspended in 500µl TE(7.5) and then extracted with Phenol/Chloroform, before ethanol precipitation of the plasmid DNA. The DNA pellet was resuspended in 100µl TE (7.5).

Caesium Chloride Preparations: This method gave DNA of very high purity: 1. 50ml overnight culture of the clone in LB Amp; 2. added to 1 litre of LB Amp and grown for a further 2 hours at 30 or 37°C or

until the optical density at 595 n.m. (A_{595nm}) was >0.5; 3. added 300m.g. of spectinomycin powder and left shaking overnight; 4. cells centrifuged at 5000g for 10 minutes at 4°C; 5. resuspended in 6.6ml 25% sucrose; 50mMTris pH8; 50mM EDTA. Added 2ml 10mg/ml lysozyme in Tris pH8; 6. added 13ml 1% Triton X-100; 250mM EDTA; 50mM Tris pH8. Mixed for 15 minutes on ice; 7. centrifuged at 5000g for 30 minutes at 4°C; 8. the supernatant was mixed with 2ml of 1% ethidium bromide solution and 0.975g of caesium chloride per ml was added to give a density of around 1.6g/ml; 9. the solution was divided into heat-sealable 5ml tubes for the Vti65 Beckman rotor. These were sealed and spun in the rotor and ultracentrifuged at 150,000g at 15°C for >6 hours. The gradient takes around 5 hours to form in the vertical rotor; 10. tubes were removed and the DNA bands were visualised in the gradient with UV. The lower band is the plasmid DNA and it was removed using a hypodermic syringe; 11. the DNA solution was extracted several times with an equal volume of sodium chloride-saturated isopropanol until the ethidium bromide had been removed; 12. 2 volumes of water and 6 volumes of ethanol was then added to the DNA solution and this was placed at -20°C overnight; 13. the DNA precipitatewas obtained by centrifuging at 10,000g for 30 minutes at 4°C, dried by vacuum, and resuspended in 1ml TE (pH7.5).

2.3.9 *Bal31* Digestion.

The exonuclease *Bal31* chews away both 5' and 3' strands of linear DNA from each end. A trial digest of any given linear DNA fragment gave a measure of the rate of digestion and allowed the desired conditions to be set. A typical *Bal31* digestion was: 5µl *Eco*RI-digested pJG(1) (3µg, 6 kb) + 20µl 2x *Bal31* buffer (1.2M NaCl, 40mM Tris pH 8, 25mM $CaCl_2$, 25mM $MgCl_2$, 2mM EDTA) +

15µl water + 0.5µl *Bal31* (0.5 units; Boehringer Mannheim). This was incubated at 20°C and 7µl aliquots were removed at 3 minute time points. The reaction was halted by addition of 10µl 0.2M EDTA and placing on ice. Samples were run on an agarose gel against Mwt. standards. When a reaction rate had been determined, the reaction was scaled up, and conditions set to remove the appropriate number of base pairs from the DNA.

2.3.10 Klenow End-Filling of DNA

The Klenow enzyme (or large fragment DNA polymerase1) fills in 5' overhangs of DNA with the appropriate homologous nucleotides. For example: 42µl *Nco*I-digested pJG(1) (20µg) + 5µl 10 x NT buffer (50mM $MgCl_2$, 50mM Tris 7.5) + 2µl dNTP mixture (200 µM of each of the unlabelled deoxyribonucleotide triphosphates) + 1µl Klenow enzyme (2 units; Boehringer Mannheim). Incubated at 30°C for >30 minutes. The DNA was subsequently cleaned up by Phenol/$CHCl_3$ extraction and ethanol precipitation, and resuspended in 50µl TE.

2.3.11 Radioactive End-Labelling of DNA fragments.

The DNA was digested with a restriction enzyme, e.g. 2µl midiprep DNA (1µg) + 0.3µl 10 x *Bam*HI buffer + 0.3µl 10 x *Hae*III buffer +1 unit *Bam*HI + 1 unit *Hae*III + 3µl water at 37°C for 1 hour. This was end-labelled by mixing with 6µl (0.6µl 10 x NT buffer + 0.1µl of each of the 200µM dGTP, dCTP and dTTP solutions + 0.1µl ^{32}PdATP(1µCurie) + 5µl water) and incubated at 30°C for 30 minutes. 6µl bromophenol blue loading buffer was then added, and the DNA sample loaded onto a 5% TBE polyacrylamide gel and electrophoresed with a molecular marker, e.g. plasmid pCE30 digested with *Sau3A* and end-labelled (sizes: 345, 262, 252, 134, 109, 82, 79, 78, 58, 50, 40, 22, 21, 16, 15, 12). The dye was run to

the bottom of the gel which was then dried and autoradiographed.

Materials

α-^{32}P dATP (250μCi in 25μl) supplied by Amersham Radiochemicals; 5% Polyacrylamide gel: 5ml 30% acrylamide stock (29:1 acrylamide: bis) + 6ml 10 x TBE + 49ml H_2O + 150μl 10% ammonium persulphate + 50μl TEMED.(N,N,N',N'-tetramethyl ethylenediamine.

2.3.12 Plasmid Sequencing.

This was based on the kit from Boehringer Mannheim. The procedure entailed denaturing the DNA to single strands, annealing a primer and then carrying out a chain termination synthesis reaction with the incorporation of labelled dNTP and dideoxynucleotides.

i) Denaturing the DNA: 1. 2μg of midiprep DNA was dried down under vacuum; 2. resuspended in 40μl denaturing buffer (0.2M NaOH, 0.2mM EDTA). Left for 5 minutes at room temperature; 3. 4μl of neutralising buffer added (2M ammonium acetate, pH 4.5) then add 100μl 95% ethanol at -20°C; 4. mix and freeze in liquid N_2; 5. spin 10 minutes, 4°C. Wash pellet with 70% ethanol. Centrifuge 1 minute. Pellet dried under vacuum.

ii) Annealing: Dried denatured DNA + 0.5μl primer solution (2.5 pmol) + 1.5μl 10 x reaction bufer + 1μl α-^{32}PdATP (10μCi) + 12μl H_2O. Incubated at 37°C for 15 minutes.

iii) Chain Elongation: 1. Annealing mixture + 1μl Klenow enzyme (2 units); 2. 3μl of this mixture pipetted into four tubes, A,G,C,T, each containing 2μl of the relevant dNTP/ddNTP mixture. Incubate at 37°C for 15 minutes; 3. reaction chased by addition of 1.5μl dNTPs (2mM); 4. terminate the reaction by adding 4μl formamide loading buffer; 5. immediately before loading the samples on the sequencing gel, denature them at 45°C for 3 minutes, then load 2μl

of each tube in a separate lane.

iv) Gel Electrophoresis: The gel was cast in the BioRad sequencing gel apparatus. Gel mixture: 10ml 40% acrylamide (29:1 acrylamide : bis) + 25.2g urea + 6ml 10 x TBE + 20ml H_2O + 165µl 10% APS + 50µl TEMED. The gel was placed in the apparatus and electrophoresed in TBE buffer. It was pre-run at 1600 volts for an hour until upto running temperature (65°C). A shark-tooth comb was then introduced, and the denatured samples loaded. The gel was run at 1600 volts until the marker dye neared the bottom. The gel was then stuck onto a sheet of 3MM paper, covered in Gladwrap, and dried under vacuum at 80°C in the BioRad gel-drying apparatus. The dried gel was placed in a lightproof cassette with a sheet of Kodak XAR-5 film, and left to expose overnight. The film was developed as directed.

CHAPTER 3.

Expression and Purification of the *rough* Homeodomain.

3.1 Introduction.

This chapter is concerned with use of the plasmids described in Chapter 2 to obtain expression of the *rough* homeodomain in *Escherichia coli*. The aim of this work was to obtain sufficiently large quantities of the *rough* homeodomain to permit its use in structural and functional studies (for 2-D nmr, a minimum of ~9 mg (1 μmol) of the protein is required (Otting et al.1988)). A variety of other genes have been expressed in essentially the same vector system, and the resulting proteins account for >10% of the bacterial cell protein, e.g. *E. coli dnaB*, *hmp*, and *ssb* (Elvin et al.1990).

Factors affecting the actual expression of the gene at the level of transcription and translation were discussed earlier. Here are considered factors that directly affect the newly-synthesized protein, and the effects that these may have upon the yield of protein folded in its native conformation. Problems associated with isolation and purification of the overexpressed protein will also be discussed.

The first problem associated with the expression of a foreign gene in *E. coli*, especially a small portion of a eukaryotic gene, is one of correct folding of the polypeptide product.

The polypeptide chain produced by the ribosomes translating the codon sequence of the mRNA will tend to adopt a particular spatial arrangement according to the interactions of the sidegroups of the component amino acids (Jaenicke 1987). However, although it has long been assumed that all of the information necessary for protein folding is encoded in the amino acid sequence, recent work has suggested that this may not always

be so (Golonbinoff et al.1989). The folding process may require specific factors, for example, chaperonins (Golonbinoff et al.1989) which are not found in the *E. coli* cytoplasm. In addition, many eukaryotic proteins require some post-translational processing and modification in order to acquire a final functional form. Even in the absence of such a processing requirement, the expression of only a portion of the native gene to give a part of the protein may result in a polypeptide which lacks the relevant amino acid residue interactions necessary for its proper folding. As such, the peptide may not be capable of adopting the configuration that it would possess in the native protein.

In the case of the homeodomain expressed from plasmid pJG(2) (Figure 2.8), the predicted sequence of 70 amino acids (shown in Figure 3.1) consists of 69 amino acids encoded by the middle exon of the *rough* gene (Tomlinson et al.1988; Figure 1.1) starting with a methionine instead of an arginine residue. This represents only 20% of the predicted full-length native *rough* protein (350 amino acids;Tomlinson et al.1988). However, since the 60 amino acids of the homeodomain seem to be very highly conserved throughout the animal kingdom despite the high degree of variability in the flanking regions, then it seemed possible that the homeodomain could be an entity capable in its own right of folding independently of the bulk of the remainder of the protein. This seems to be borne out by the successful purification of an apparently functional *Antennapedia* homeodomain (Muller et al.1988; see Figure 6.1) which consists of 67 amino acids (or 18% of the native *Antennapedia* protein; Schneuwly et al.1986).

The protein produced may be subject to extensive degradation by the proteases of the bacterial cell. *E. coli* possesses a series of proteases which display a variable specificity for, and rate of degradation of, abnormal proteins (Goldberg and Goff 1986). It is noted that small proteins are especially susceptible to such

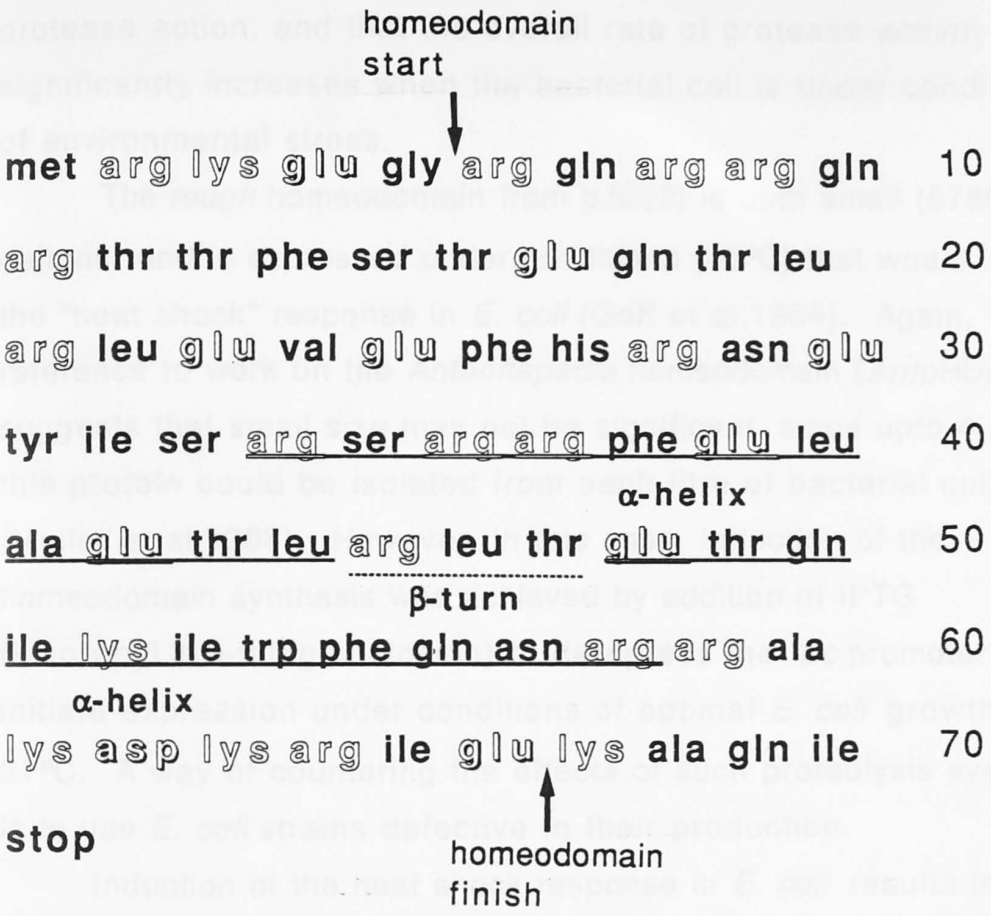


Figure 3.1: The 70 amino acid sequence of the *rough* homeodomain protein encoded by pJG(2).

The start and finish of the 61 residue *rough* homeodomain are shown. There are another 4 residues encoded by the *rough* gene at either end of the homeodomain; the met-1 is the only amino acid not encoded by *rough* in this position. The putative helix-turn-helix motif within the homeodomain is indicated (see Section 4.1.3). This protein has a high proportion of charged residues with 18 basic (arg/lys) and 10 acidic (glu/asp) side chains, giving the protein a strongly basic charge overall. The molecular weight of the protein obtained by summation of the residues is 8786 daltons.

protease action, and that the overall rate of protease activity significantly increases when the bacterial cell is under conditions of environmental stress.

The *rough* homeodomain from pJG(2) is both small (8786 daltons) and is expressed under conditions (42°C) that would induce the "heat shock" response in *E. coli* (Goff et al.1984). Again, reference to work on the *Antennapedia* homeodomain (*Antp*HD) suggests that small size may not be significant, since upto 4 mg of this protein could be isolated from each litre of bacterial culture (Muller et al.1988). However, in this case, induction of the homeodomain synthesis was achieved by addition of IPTG (isopropyl β-D-thiogalactoside) to derepress the *lac* promoter and initiate expression under conditions of optimal *E. coli* growth at 37°C. A way of countering the effects of such proteolysis systems is to use *E. coli* strains defective in their production.

Induction of the heat shock response in *E. coli* results in a doubling of the concentration of the ATP-dependent protease La (the *lon* gene product) and a consequent increase in the rate of proteolysis of abnormal proteins (Goff et al.1984). It is suggested that La catalyses the rate-limiting step in the breakdown of most abnormal proteins (Goldberg and Goff 1986). Since mutations in *lon* result in a generalised defect in the degradation of abnormal proteins, a *lon E. coli* strain might be used as the host for an expression system.

Mutations in the *htpR* gene also affect the normal functioning of the proteolysis system in *E. coli* (Goff et al.1984). The *htpR* gene encodes an alternative σ–subunit of RNA polymerase necessary for transcription of the heat shock genes and hence, for induction of the heat shock response. One aspect of its function is enhancement of protease La activity. It seems that even under conditions of normal growth (e.g. 30° or 37°C), *htpR* mutants

display a lower rate of degradation of many proteins than do wild-type cells (Goff et al.1984). Thus, a double *htpR-lon* mutant may be an even better host strain if proteolysis is a problem.

Another feature associated with the high-level expression of cloned genes is aggregation of the overproduced protein (Buell and Panayotatos 1986). This process may entail misfolding of the protein (Jaenicke 1987) or its conversion from a soluble state (from which it may be recovered by extraction with neutral aqueous buffers) into an insoluble state (from which the protein may be recovered only by extraction with detergents or chaotropic agents). The aggregated protein tends to occur as dense osmiophilic inclusions ("inclusion bodies") which can be visualised in cells by electron microscopy and have been found in some cases to account for 20% of the cellular volume (Williams et al.1982). Although, these structures were initially thought to represent some sort of degrading organelle in which abnormal proteins might accumulate prior to degradation (Goldberg and Goff 1986), they are not bounded by membrane and do not contain proteolytic enzymes. Instead, they appear to be amorphous aggregates of partially or completely denatured protein. The formation of such structures seems to be a spontaneous process occurring when cells produce large amounts of foreign protein. They are generated from soluble proteins even when ATP production ceases (Goldberg and Goff 1986). However, the subsequent removal of proteins from such aggregates requires ATP and involves the complete hydrolysis of the proteins (protease La requires ATP). Hence, it has been suggested that the formation of protein aggregates may be part of the cell's degradation process, e.g. as a physical consequence of the saturation of its degrading capacity, or in order to direct the proteolytic digestion of foreign proteins.

It has been observed that aggregate pellets of the overexpressed polypeptides, easily isolated by differential

centrifugation, are quite uniform in composition. After prolonged expression of a cloned gene, they often have a rather homogeneous, almost crystalline appearance (Williams et al.1982).

The tendency of unfolded foreign proteins to form dense intracellular aggregates varies widely according to the protein, and the host *E. coli* strain used (Buell and Panayotatos 1986). It has, however, proved useful in isolation of the denatured protein, because it allows the selective removal of soluble contaminating bacterial proteins. Indeed, this technique is precisely that used for isolation of the *rough* - β-galactosidase fusion proteins, whose production is discussed in Chapter 5. The apparent tendency of the *rough* homeodomain to form such insoluble aggregates is discussed in Section 3.2.

This now leads to a consideration of the purification procedure to be used. The predicted *rough* homeodomain is small, highly charged, and very basic (it has 19 basic to 10 acidic amino acids, giving an estimated isoelectric point >pH10). If it were to form intracellular aggregates (as described above), it would be necessary to develop a procedure for its resolubilisation before chromatographic methods could be utilised.

Comparison with the published procedure for isolation of the *Antennapedia* homeodomain (Muller et al.1988) might be instructive. The presence of the *Antennapedia* homeodomain (*Antp* HD) was monitored during its purification by visualising it as a band on SDS-polyacrylamide electrophoresis gels. The procedure began with the lysis of the induced cells in a French Press. The *Antp* HD was in the soluble fraction obtained by centrifugation of the lysis mixture. Polyethylene imine (Polymin P), a polyamine polycation, was added to this fraction to precipitate out remaining DNA fragments. This was done at high salt concentration (0.4 M NaCl) presumably because it was thought that a basic protein like *Antp* HD (20 basic versus 7 acidic residues) might bind anionic DNA

fragments, and might otherwise be precipitated with them. The Polymin P-treated mixture was centrifuged and the supernatant was loaded onto a cation-exchange chromatography column. The matrix used, Bio-Rex A70, carries carboxyl groups (-COOH; pK_a~5) and so at neutral or higher pH values bind cationic proteins. The bound proteins are removed by elution with increasing concentrations of a competing cation (e.g. Na^+). At the appropriate [Na^+], the *Antp* HD was displaced and collected. It may not have been pure since, following dilution of the fractions to lower [Na^+], they were applied onto another cation-exchange resin (Mono-S [Pharmacia]: the active goup is -SO_3^-). Elution of the bound proteins from this column with an increasing [Na^+] gradient gave fractions in which the *Antp* HD was pure.

By contrast, in this work, it was not possible initially to detect the *rough* homeodomain as a readily identifiable band on SDS-polyacrylamide gel electrophoresis (SDS-PAGE) gels. Neither was there an assay for the presence of the protein in fractions, although there were hopes of using the polyclonal antibodies raised against the X17 fusion protein (see Chapter 5) to detect it immunochemically. In the absence of information on the specificity of DNA binding, it was impossible to contemplate a DNA-binding assay.

The first objective, therefore, was to develop a procedure for isolation of small quantities of the *rough* homeodomain (*ro*HD), irrespective of yield, of sufficient purity to enable it to be used to establish a specific DNA-binding assay. The refinement of the purification procedure would then follow.

3.2 Results and Discussion

This section gives an account of the expression and

purification studies of the *rough* homeodomain (*ro*HD). Section 3.2.1 gives details of the attempts to confirm that there had been expression of the *ro*HD in induced pJG(2) cells. Although not apparently expressed to the levels expected for this expression system (Elvin et al.1990), the *ro*HD was visualised on SDS-PAGE protein gels, but was found to be migrating at an anomalously high relative molecular weight. It was shown to be present in an insoluble, presumably aggregate form, although there was evidence that it was also present in the soluble form predicted for such a globular protein. Section 3.2.2 describes the attempts made to purify the soluble *ro*HD using a similar procedure to that successfully used to purify the *Antp* HD (Muller et al.1988). Though this did result in the purification of the *ro*HD, the yield was very low. Section 3.2.3 is an account of other purification methods which were tried to improve the yield of the *ro*HD. Finally, Section 2.3.4 describes the successful purification of 0.6 mg of the *ro*HD by extraction of the insoluble protein with urea solution, and its renaturation, before ion-exchange chromatography at high pH. Further purification was not attempted until an assay based upon the DNA-binding properties of the *ro*HD could be developed (see Chapter 4).

3.2.1 Initial Expression Work.

The first problem in the isolation of the *ro*HD was to identify it in cell extracts.

The initial expression studies were conducted with the 5' plasmids: pJG(1)10, pJG(1)13 and pJG(1)C11, whose construction, and subsequent incorporation into the final expression construct, is described in the preceding chapter. These studies involved the thermal induction (at 42°C) of small broth cultures of exponential phase cells. Treatment at 42°C proceeded for variable periods

from 30 minutes to fifteen hours. Cells were harvested by centrifugation and lysed in 1.2% sodium dodecyl sulphate (SDS) at 100°C. This procedure solubilises most of the bacterial proteins. SDS is a detergent that binds to polypeptides (at a weight ratio of 1.4 :1, SDS: protein; Berger and Kimmel 1987), and dithiothreitol is used to reduce disulphide bonds. The protein solution is loaded onto a polyacrylamide gel (containing 0.1% SDS) and separated by electrophoresis (SDS-PAGE), the SDS-protein complexes are negatively-charged and migrate towards the anode (Laemmli 1970). Comparison of the protein composition of these SDS cell lysates before and after treatment at 42°C, electrophoresed against molecular weight standards, gives a clear indication of accumulation of the desired protein, provided it is produced to levels in excess of around 0.5% of total protein in the cell (Elvin et al. 1990). Bacterial heat-shock proteins are also readily identified. There was, however, no significant change apparent in the composition of induced SDS lysates from the 5' plasmids (data not shown). The *rough* open reading frame in these plasmids should have given a protein with a molecular weight between 12,700 and 13,300 daltons, and the gels were run to allow a good separation of bands in this region. If there was an additional protein band present in the induced SDS lysate, it was either unexpectedly weak, and/or was being obscured by another protein band, present in lysates of the induced and uninduced lanes.

It was hoped that this apparently poor yield might be associated with instability of the *rough* protein product due to the 35 amino acids at the 3' end which were removed in the final homeodomain constructs, pJG(2) and pJG(3). Hence, once constructed, the expression work concentrated on the plasmid pJG(2) .

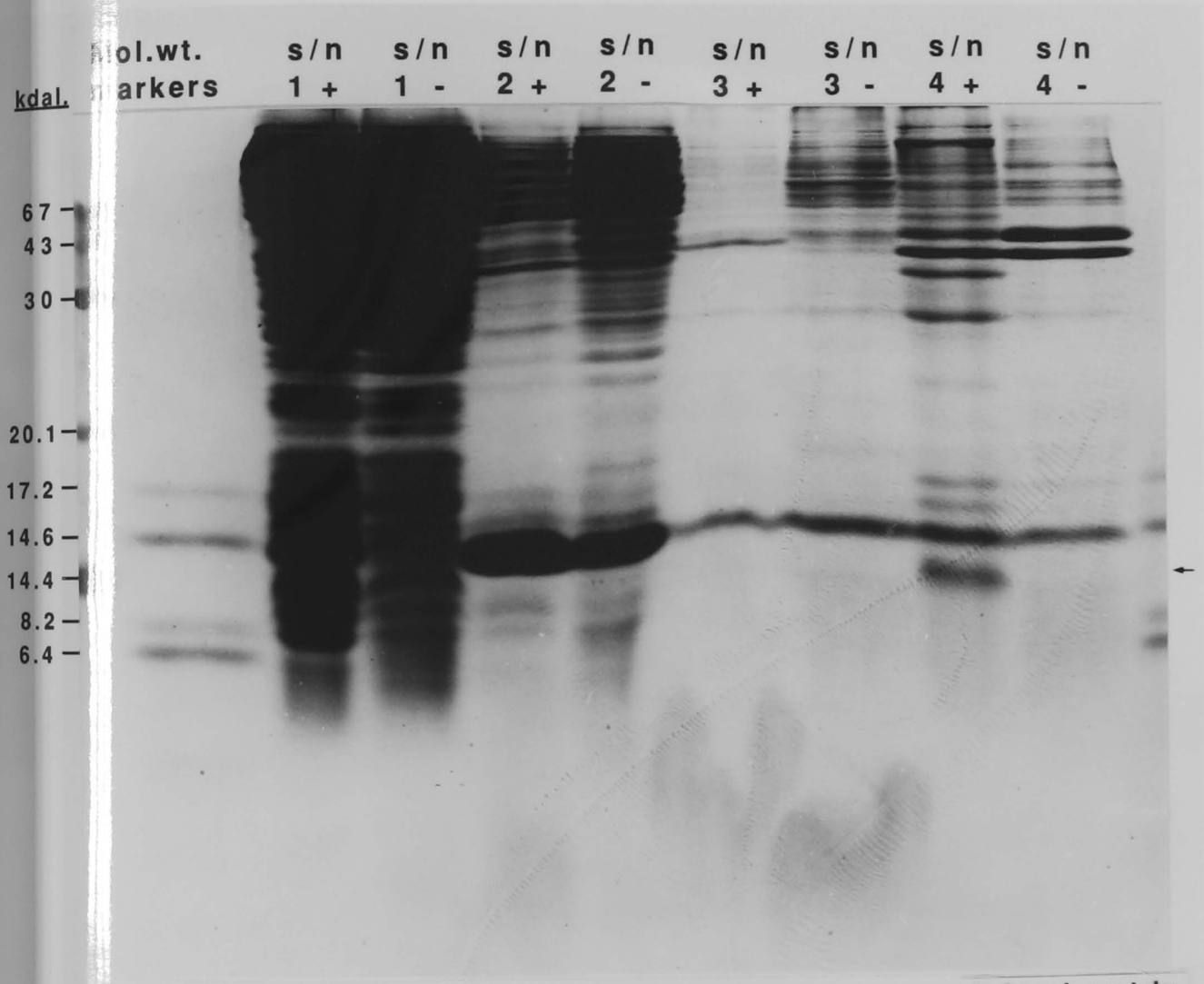
From the predicted amino acid sequence of the *ro*HD encoded

by pJG(2) (given in Figure 3.1), the molecular weight was calculated to be 8786. SDS lysates of an induced culture of cells containing pJG(2) were prepared, and run on SDS-PAGE. However, there were still no differences evident between the induced and uninduced lanes in this region (Figure 3.5).

Given that a strain containing pJG(2) is clearly temperature sensitive in its growth, there would seem to be a degree of activity associated with the temperature-inducible expression system. Hence, it was reasoned that the homeodomain was being produced in the induced clones but that, for some reason possibly associated with its small size and high positive charge, its presence was not evident on SDS-PAGE gels (Weber and Osborn 1975).

Two main possibilities were considered at this juncture. Either the protein was present in the lysate and entering the gel but was not evident due to inadequate staining or fixing in the gel, or that it was not present in the lysate because it was in the material not solubilised by the SDS treatment. Aggregation was found to occur with the fusion protein (discussed in Chapter 5). Indeed, the fusion protein was not readily solubilised in SDS at 100°C, and did not enter the SDS-polyacrylamide gel under the conditions used here.

To test the possibility that the protein was entering these gels but was not evident for reasons associated with the nature of the protein separation in the gels, or due to the gel staining protocol used, variations in the composition of the polyacrylamide gels were tried, with gradient gels, higher and lower percentages of acrylamide monomer in the separating gel, the addition of glycerol to the gel, stronger fixation procedures, and stronger Coomassie blue staining solutions (Giulian et al.1984; DeWald et al.1986). None of these measures resulted in clear visualisation of



Induction: pJG(2) cells were grown at 30°C to A_{595nm} ~0.5; they were then grown at 42°C for 2 hours, and harvested.

Extraction procedure: the cell pellet was r/s in sonication buffer (50mM Tris pH 7.5, 10% sucrose). This was sonicated for 1 minute on ice and centrifuged at 12,000 g (supernatant = s/n 1). The pellet was r/s in T.N.E. (25mM Tris pH 8, 50mM NaCl, 2mM DTT, 10% sucrose) + 100µg/ml lysozyme for 30 minutes on ice, then frozen in liquid N_2, thawed, and centrifuged at 12,000 g (s/n 2). The pellet was r/s in T.N.E. + 10% Triton X-100 for 10 minutes on ice, and centrifuged (s/n 3). This pellet was r/s in 8M urea (10mM Tris 7.4) for 1 hour on ice, and centrifuged (s/n 4).

Figure 3.2 : SDS-polyacrylamide gel of protein extracts from pJG(2) cells grown at 42°C (+) and 30°C (-).

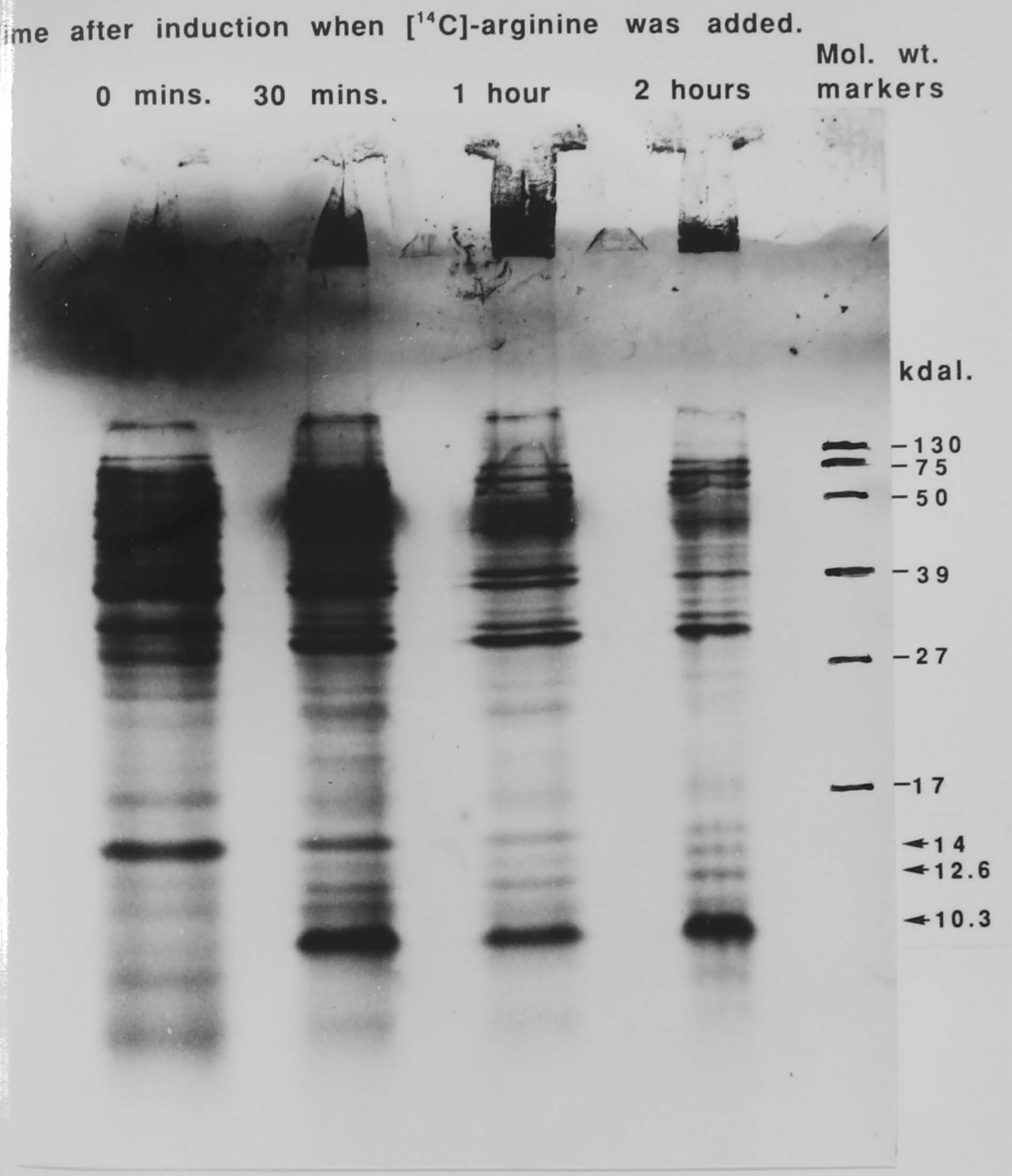
the homeodomain on SDS-PAGE gels.

The method used for solubilising the fusion protein (see Chapter 5; Howell and Hargreaves 1988) was tried on induced and uninduced cultures of cells containing pJG(2) (Figure 3.2). Cells were first lysed by sonication, and were then centrifuged to remove soluble proteins (s/n 1). The pellet was resuspended and other proteins were solubilised by freezing and thawing in the presence of lysozyme. The soluble fraction (s/n 2) was again removed by centrifugation. The pellet was resuspended in 10%(w/v) Triton X-100. Following centrifugation (s/n 3), the pellet was finally extracted several times with 8 M urea (s/n 4).

Soluble fractions for each of these four steps were treated with SDS, and run on SDS-PAGE gels. A difference between uninduced and induced cells was now apparent (see Figure 3.2).

In the first supernatant (s/n 1), there appeared to be an additional protein (around 15 kdal) in the induced lane. However, it was in the final urea extract (s/n 4) that the clearest difference was apparent, with the presence in the induced lane of a protein around 14 kdal. Thus, there was indeed an insoluble protein whose presence seemed to be restricted to the induced cells. Yet, given that the expected molecular weight of the *rough* protein was 8,800, the identity of this protein was not clear. The anomalous migration of small basic proteins on SDS gels, to give higher apparent molecular weights, has however been documented frequently (e.g. histones (Panyim and Chalkley 1971), HU (Rouviere-Yaniv and Gros 1975)). In order to clarify the situation, an attempt was made to identify induced proteins by the incorporation of a radioactively-labelled amino acid during induction. Since the *rough* homeodomain (*ro*HD) was predicted to contain a large number of arginine residues (14 of 70), [^{14}C]-arginine was used.

The incorporation procedure entailed growing small liquid



Method: pJG(2) cells grown to A_{595nm} ~0.5 in LB broth at 30°C, cells harvested at 5000g, r/s in minimal media, and then induced by growth at 42°C. Addition of 2.5 μCi/ml [^{14}C]-arginine at 0, 30, 60, and 120 minutes after induction. Cells harvested 30 minutes after addition of arginine, and r/s in SDS-lysis buffer; electrophoresis through 20% separating gel.

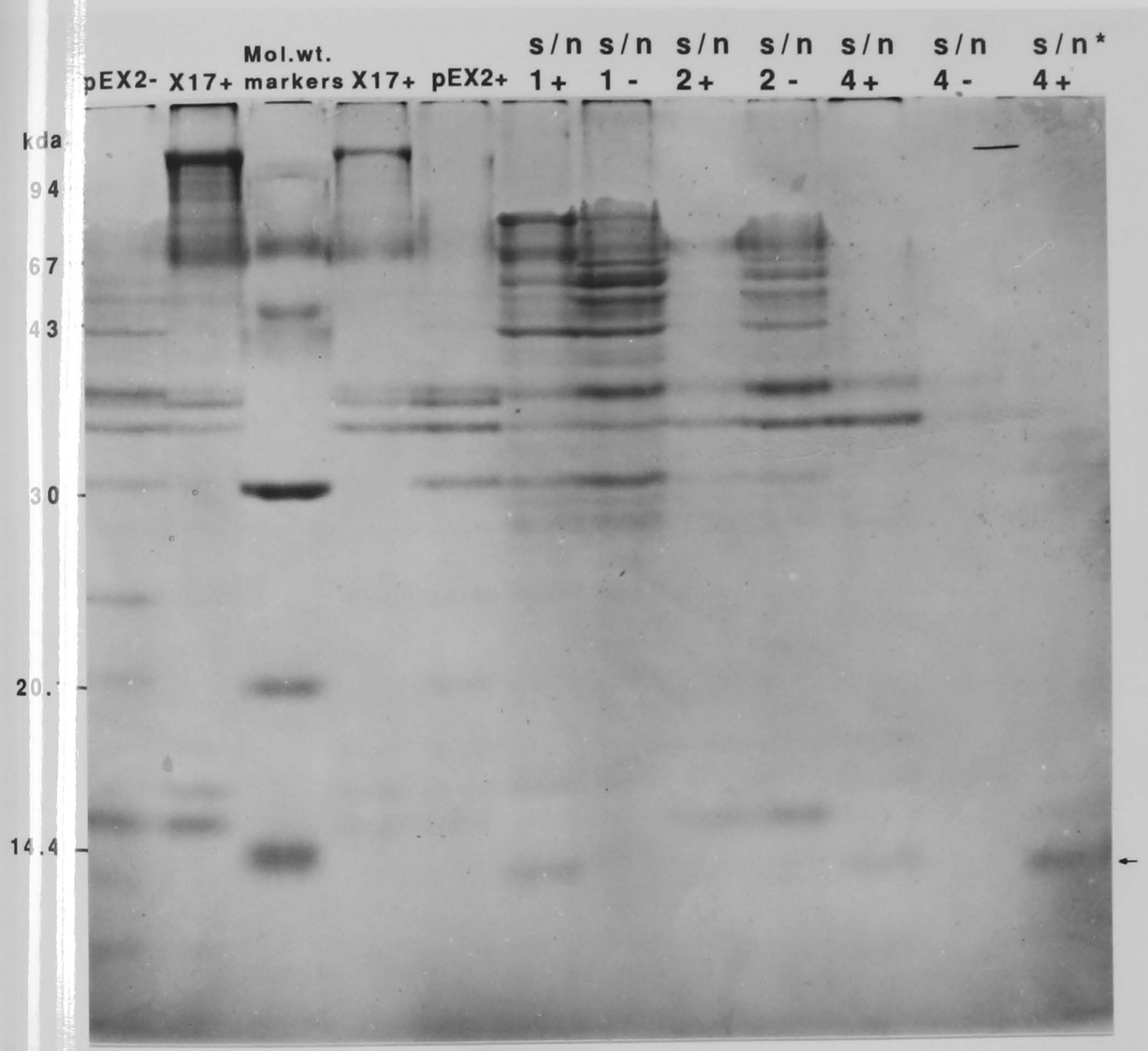
Figure 3.3: Autoradiograph of SDS-polyacrylamide gel showing incorporation of [^{14}C]-arginine into induced and uninduced protein extracts of pJG(2).

cultures of pJG(2) at 30°C in Luria-Bertani broth with ampicillin (LBAmp; see Methods) until they were in the exponential growth phase. The cells were then centrifuged, and resuspended in appropriately supplemented 56-minimal medium. After a further period at 30°C, the culture was heated to 42°C and [^{14}C]-arginine was added. Samples were taken after 0.5, 1 and 2 hours. These were lysed with SDS, and separated by SDS-PAGE. The labelled proteins were visualised by autoradiography. There were three novel protein bands present in induced cells in comparison with uninduced cells (see Figure 3.3). The strongest of these had a molecular weight of 10.3 kdal, the other two were at 12.6 and 14 kdal. It is presumably the latter band which corresponds to the protein apparent in the urea extract (s/n 4+; Figure 3.2).

Hence, the incorporation study suggested that there was indeed a strongly induced protein that had close to the predicted molecular weight of the *ro*HD yet, despite the clear band on this gel (possibly due to the abundance of arginine), there was no correspondingly outstanding band of this size on Coomassie-blue stained SDS-PAGE gels (Figure 3.5). Rather, the lowest band showing a significant change migrates around 14 kdal.

Attempts were made to detect the *ro*HD using the polyclonal antibodies raised against the X17 fusion protein (Chapter 5). Since the X17 protein contains most of the *ro*HD (there are 58 amino acids in common), the antisera should display some specificity for the protein.

Western blots of SDS-lysates of induced cells containing pJG(1)10(4), pJG(1)C11 and pJG(1)13 showed an antigenic band at around 22 kdal, but no other noticeable differences with the uninduced lysates. A Western blot of the pJG(2) extracts (s/n's 1, 2,and 4; Figure 3.2) did not identify any of the proteins shown in the ^{14}C-incorporation studies, and background staining of other



* dialysed, but not immunoprecipitated.
← *ro*HD in s/n 4+

Protein extracts: pJG(2) extracts as in Figure 3.2, s/n 4 +/- were dialysed against T.N.E. before immunoprecipitation. pEX2- = pEX2 in N4830 grown at 30°C and prepared as for pJG(2) s/n 1-. pEX2+ and X17+ = induced extracts prepared as for pJG(2) s/n 4+. Details of pEX2 and X17 are given in Chapter 5.

Immunoprecipitation procedure: samples of the protein extracts were mixed with antisera (1:500 dilution) raised against the X17 fusion protein (see Chapter 5 for details) in T.N.E. buffer. After 72 hours at 4°C, the mixtures were centrifuged at 16,000g for 10 minutes. The pellets were resuspended in SDS-lysis buffer and electrophoresed on a 13% SDS-polyacrylamide gel.

Figure 3.4: SDS-PAGE showing immunoprecipitation of proteins from pJG(2) extracts (see Figure 3.2) and pEX2 / X17 extracts (see Chapter 5).

proteins, notably lysozyme in the freeze/thaw supernatant, was evident (not shown).

Immunoprecipitation was tried as an alternative method to the Western blot since, as a result of staining the SDS-polyacrylamide gel after it had been blotted, it was evident that only a partial transfer of proteins (other than the pre-stained markers) to the nitrocellulose membrane was being achieved (see Section 5.5). Immunoprecipitation with the X17 antisera (see Figure 3.4) of the proteins in the pJG(2) extracts (s/n's 1, 2, and 4; Figure 3.2) showed that the main difference was in s/n 1+ and s/n 4+ with a protein at ~14 kdal. It would appear to be the same protein immunoprecipitated from both extracts. This protein was already evident in s/n 4+ (Figure 3.2), but was not clear in s/n 1+. Thus, it appeared that this protein occurred in the cell in both an initially soluble form, readily extracted from induced pJG(2) by gentle lysis, and in an insoluble (presumably aggregated) form, which was not extracted by sonication or treatment with 10% Triton X-100 (s/n 2+, s/n 3+), but was finally extracted by 8 M urea (s/n 4+).

Given the result with the immunoprecipitation, it was thought that perhaps this 14 kdal protein might indeed have been the *ro*HD, despite the [^{14}C]-arginine incorporation result with its more promising band. Ultimately, however, the only way to satisfactorily establish its identity was by determination of the N-terminal sequence of the protein.

The s/n 4+ urea extract from pJG(2) (Figure 3.2) was transferred electrophoretically from an SDS gel onto a polyvinylidene difluoride (PVDF) membrane (Matsudaira 1987). The membrane was stained with Coomassie blue to visualise proteins. The 14 kdal band described above was excised and the N-terminal sequence was determined by Dr D.Shaw (John Curtin School of

Figure 3.5: SDS-PAGE showing the effect of time of induction on pJG(2) cell protein.

pJG(2) cells were grown at 30°C until A_{595nm}= 0.53; they were then grown at 42°C to induce expression of the *ro*HD. Samples were removed from the induced culture at 0, 0.5, 1, 2, 3, and 15 hours, their A_{595nm} measured, and equal aliquots solubilised with SDS buffer, before loading on 20% SDS-PAGE.

The protein region around 14 kdal is indicated by the solid bar. There seems to be an increase in the relative protein concentration in this region with time of induction, reaching a plateau around 2-3 hours. The rate of increase in the density of the pJG(2) culture at 42°C declines relative to its growth rate at 30°C; pJG(2) growing exponentially at 30°C reaches its growth plateau (A_{595nm}~1.9 in LB-Amp broth) ~2-3 hours after A_{595nm}= 0.5; however, cells induced at this stage reach a much lower plateau within 3-4 hours (A_{595nm}~1.3), and do not achieve the same density even after 15 hours of growth at 42°C.

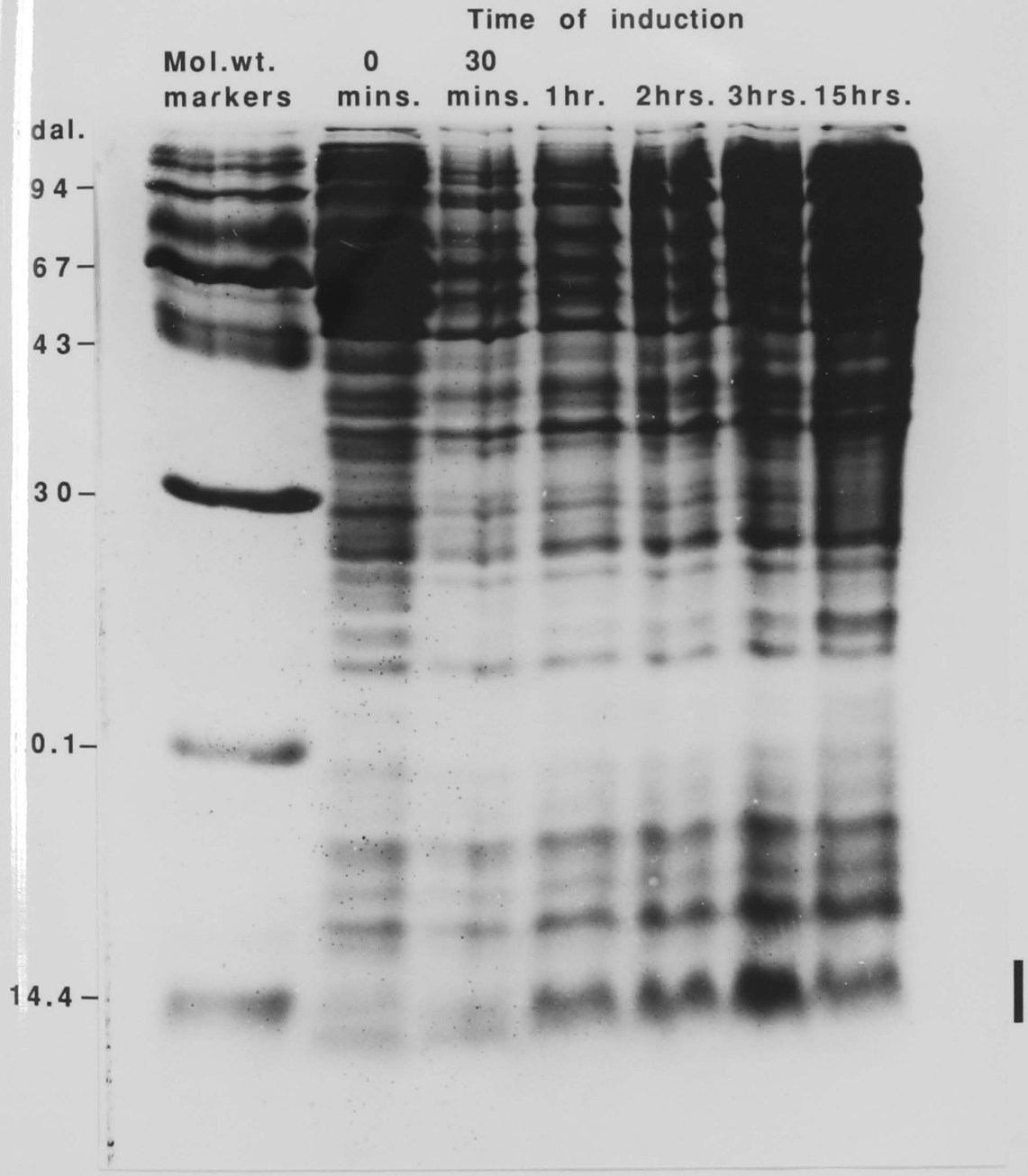


Figure 3.5: SDS-PAGE showing the effect of time of induction on pJG(2) cell protein.

Medical Research, ANU). A sequence for the first 15 amino acids was obtained and was found to match that expected for the *rough* homeodomain (Figure 3.1).

It is therefore apparent that the *ro*HD was being expressed from pJG(2), but not at as high a level as was expected based on expression work with other genes in this vector system. The question remains: why is its apparent mobility on SDS-PAGE around 14,000 daltons given its predicted molecular weight of 8,800. One possibility is that the protein formed a dimeric structure that was not easily denatured. However, other small highly basic proteins have been found to run at anomalously high molecular weights (perhaps due to a cancelling of some of the overall negative charge which dictates its rate of migration in SDS-PAGE) (Panyim and Chalkley 1971; Rouviere-Yaniv and Gros 1975; Weber and Osborn 1975).

Thus, this work suggested that the *ro*HD was distributed between two fractions - the insoluble portion seemed to be the more predominant, although at least some of it was soluble.

3.2.2 The *Antennapedia* Protocol.

Muller et al.(1988) reported production and purification of the *Antennapedia* homeodomain (*Antp*HD). An outline of the purification procedure for this protein was presented in the Introduction to this Chapter. Given the similarities between the *Antp* HD and *ro*HD (see Figure 6.1), the published procedure was initially followed. With the apparently low yield of *ro*HD, much larger quantities of cells were required. Time courses of the induction (Figure 3.5) suggested that production of the 14 kdal protein increased beyond 2 hours and reached a constant level that was maintained beyond 15 hours. A 40-litre culture in a batch fermenter was grown to $O.D._{595}$ ~0.8, then treated at 42°C for 4

Figure 3.6: SDS-PAGE and elution profile of the pJG(2) Polymin P supernatant fractions from a BioRex A-70 cation-exchange column.

Preparation of the protein extract: Induced pJG(2) cells (2.5 g net weight) in 30 ml of lysis buffer (Buffer A, 10% sucrose, 2 mM EDTA, 0.4 M NaCl) were lysed at 10,000 p.s.i. The supernatant from centrifugation at 12,000g (15 minutes) was taken to 0.8% Polymin P, mixed for 15 minutes, and then re-centrifuged. This supernatant was applied to, and eluted from, a BioRex column, removing DNA and Polymin P from the protein, as described in Section 3.2.2. The protein fraction from this column (= Polymin P s/n) was re-applied to a BioRex column.

BioRex chromatography: The 5 x 0.7 cm BioRex Á-70 column was pre-equilibrated with 50 ml Buffer A, 0.1 M NaCl. After the protein extract (above) had been applied, the column was washed with 100 ml Buffer A, 0.1 M NaCl (flow-through = protein washed from column at this stage). The column was eluted with a 40 ml NaCl gradient (0.1 to 1.1 M NaCl in Buffer A). 1 ml fractions were collected and their A_{280nm} determined. The graph shows a plot of the elution profile with increasing [NaCl]. 20 µl samples of elution fractions were mixed with an equal volume of SDS-lysis buffer, and electrophoresed on an 18% SDS-polyacrylamide gel as described in Section 3.3.1.

→ = the protein in fractions 8-11; * = the protein in fractions 15/16. Further purification of these proteins by Mono-S chromatography is shown in Figure 3.7.

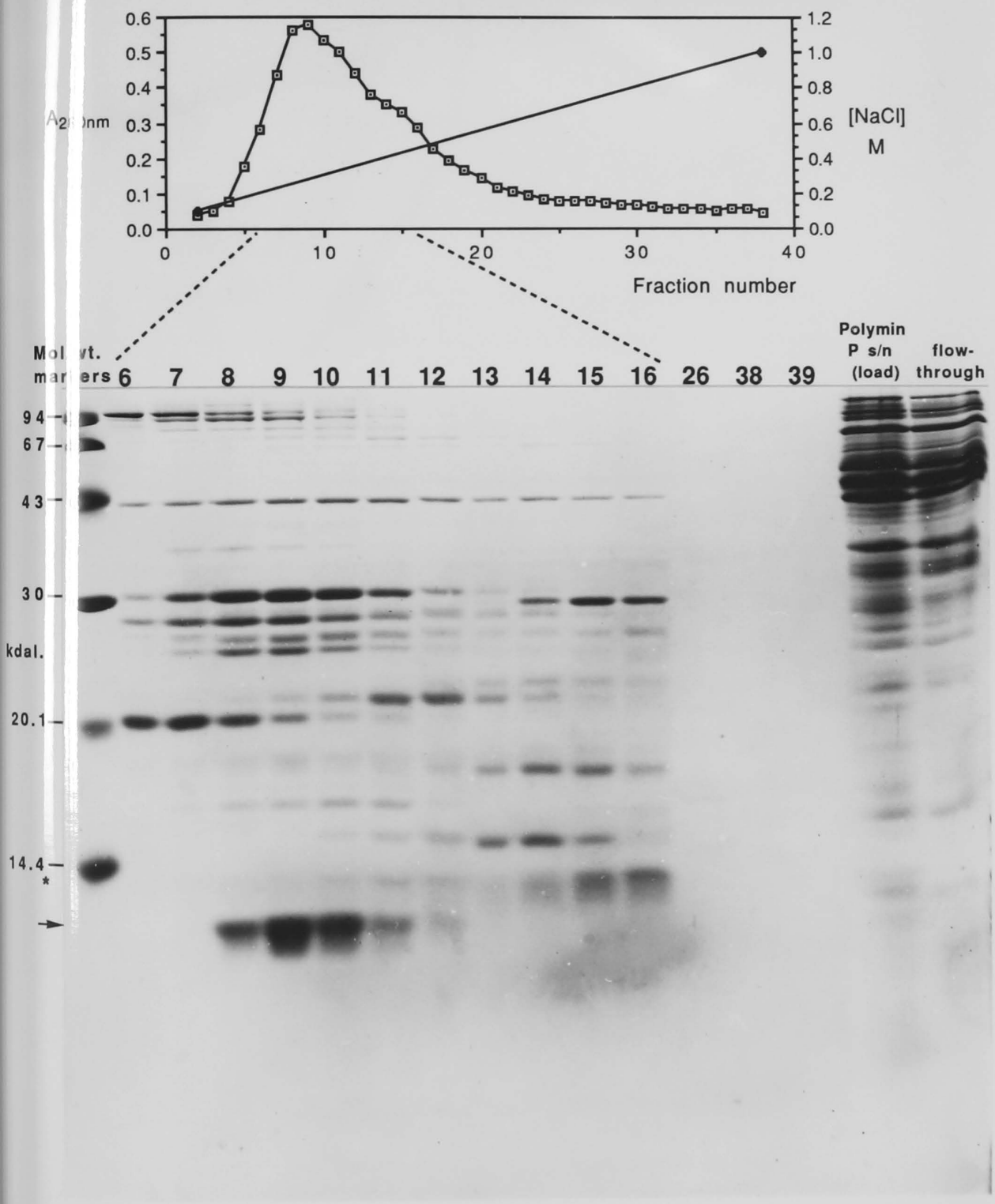


Figure 3.6: SDS-PAGE and elution profile of the pJG(2) Polymin P supernatant fractions from a BioRex A-70 cation-exchange column.

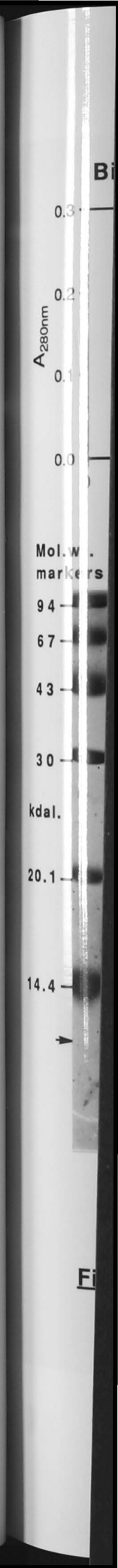
hours before the cells were harvested. Cells (141g net weight) were resuspended with an equal volume of buffer to give a slurry which was frozen in liquid nitrogen and stored at -70°C.

A sample of these cells was thawed and lysed using a French Press (at 10,000 p.s.i.). The resulting lysate was centrifuged (12,000g, 15 mins) and the supernatant was treated by dropwise addition of Polymin P to a concentration of 0.8% (v/v). This was mixed for 15 minutes and centrifuged (12,000g, 15 mins.). Following Muller et al.(1988), the supernatant was loaded directly onto a BioRex A-70 cation-exchange column (5 x 0.7 cm) pre-equilibrated with 0.4 M NaCl in Buffer A (50mM phosphate, pH 7.5). The resin was thoroughly washed and bound protein was eluted with a 40-ml gradient of 0.5 M to 0.9 M NaCl in Buffer A. The absorbance at 280 nm (A_{280}) was measured for the resulting 1-ml fractions, and samples of those with an absorbance >0.1 were loaded onto an SDS-PAGE gel (18%). Very little protein, and no protein <17 kdal, was detected (not shown). The only difference on SDS-PAGE between the extract loaded onto the column and that which flowed through, was the loss of the streaking due to Polymin P. The spectrum of eluted fractions with a high A_{280} but no apparent protein showed an absorbance maximum at 260 nm, and their fluorescence with ethidium bromide confirmed a high concentration of DNA. Since hardly any protein had been removed from the initial supernatant, but it appeared that most of the Polymin P and DNA had been, the flowthrough fraction was reapplied to another BioRex column (5 x 0.7 cm). This new column was pre-equilibrated with 0.1 M NaCl in Buffer A, and the protein solution was diluted to 0.1 M NaCl before being applied to the column. The column was washed and then eluted with a 0.1 M to 1.1 M NaCl gradient to give forty 1-ml fractions. The A_{280} of these fractions showed there to be a large peak from fraction 5 to 24.

Figure 3.7: SDS-PAGE and elution profile from Mono-S cation-exchange columns loaded with BioRex elution fractions in Figure 3.6.

Protein extract: Fractions 8-11 and Fractions 15/16 from Figure 3.6 were dialysed to Buffer B, 0.05 M NaCl, and then applied to the Mono-S column (=load).

Mono-S chromatography: The pre-packed 5 ml Mono-S column was used with the Pharmacia FPLC apparatus. The column was pre-equilibrated with 10 ml Buffer B, 0.05 M NaCl. The above protein extracts were then applied to the column, which was washed with 10 ml Buffer B, 0.05 M NaCl before developing the NaCl elution gradient (30 ml 0.05 to 1 M NaCl in Buffer B). The A_{280nm} of the 1 ml elution fractions is plotted against the [NaCl].
20 µl samples of elution fractions were mixed with an equal volume of SDS-lysis buffer, and electrophoresed on an 18% SDS-polyacrylamide gel.
→ = HU I/II; * = *ro*HD; determined by protein sequencing as described in Section 3.2.2.



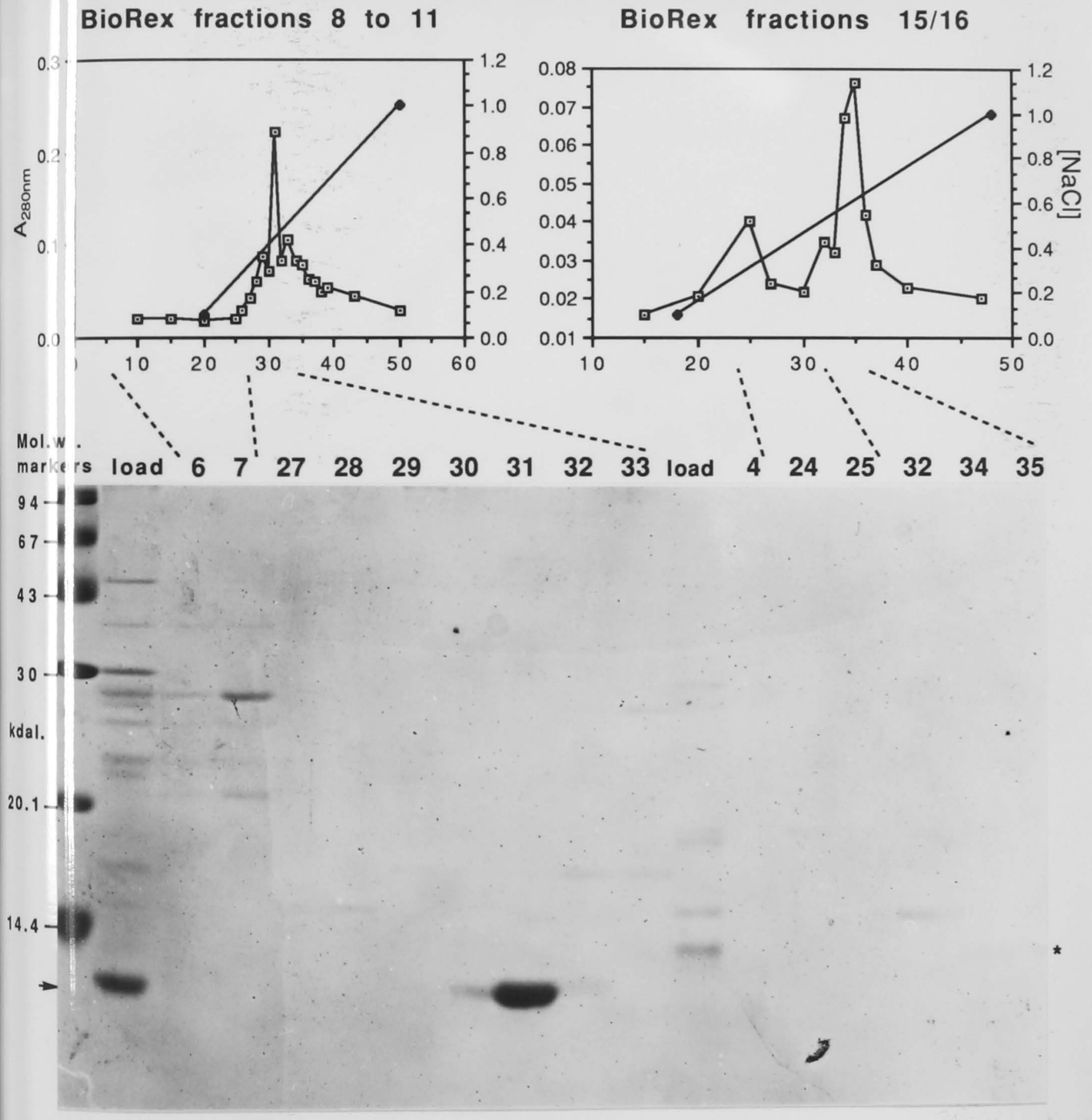


Figure 3.7: SDS-PAGE and elution profile from Mono-S cation-exchange columns loaded with BioRex elution fractions shown in Figure 3.6.

Samples of these fractions were run on an SDS-PAGE gel(Figure 3.6). Two proteins migrating below 14,400 were evident, one ~12 kdal eluted at ~0.4 M NaCl, the other, less abundant ~14 kdal, eluted at ~0.55 M NaCl.

Both of these proteins were further purified following dialysis to 50 mM NaCl in Buffer B. They were loaded separately onto a Mono-S column (5 ml), using a complete FPLC system (Pharmacia). This column was run at higher pH with Buffer B (20 mM glycine, pH 9.0) since it was reasoned that the *ro*HD would be a cation even at high pH. The column was developed with a 40-ml gradient of 50 mM to 1.0 M NaCl. The elution profile (A_{280}) of the column, and SDS-PAGE of these fractions is shown in Figure 3.7. The first protein was purified in a fraction eluting at ~0.3 M NaCl, whilst the second protein was eluted at around 0.55 M NaCl.

Both of these purified fractions were dialysed against water and the first 10 to 15 amino acids of their N-terminal sequences were determined by Dr Shaw. The more abundant protein (~12 kdal) was the small, basic *E. coli* DNA-binding protein HU I/II (Rouviere-Yaniv and Gros 1975) which occurs as a mixture of two almost identical proteins, each 90 residues long. This is an abundant protein in *E. coli* with at least 10,000 copies per cell (Rouviere-Yaniv and Gros 1975). The other protein, which was found to comigrate on SDS-PAGE gels with the 14 kdal *ro*HD identified earlier (s/n 4+), was confirmed as the *ro*HD.

Unfortunately, the yield of pure *ro*HD obtained by this means was very low. From 5g of resuspended cells (equivalent to ~750 ml of cell culture), the procedure gave ~10µg of protein (based on an assessment of the concentration of the aliquot submitted for protein sequencing, and from SDS-PAGE).

Clearly, changes in the procedure were required. However, in the absence of an assay for the protein, it was not clear just how much soluble protein was available to be purified. The

initial expression work in the last section suggested that a significant proportion of the *ro*HD was present in an initially insoluble form. Indeed, although this procedure clearly showed that the *ro*HD was also present as soluble protein in induced pJG(2) cells, it was possible that much of it was not being extracted from the lysates, due to retention of the soluble protein in the pellet at the lysis stage. This is possible since the homeodomain is expected to be fairly basic and ought to bind DNA non-specifically. Lysis could be carried out in buffer at high [NaCl]. Even then, the French Press seemed to generate a considerable quantity of smaller DNA fragments which were not initially pelletted; the protein may interact with these. As such, it may be precipitated in association with the DNA when the supernatant is treated with Polymin P. It was also possible that the protein was being degraded between the lysis and application to the column. These questions are investigated in the next section.

3.2.3 Variations on the Theme.

It was decided to try to improve the yield of soluble *ro*HD in the lysis supernatant by varying the conditions of cell lysis and the fractionation of the supernatant.

A higher [NaCl] (0.6 M) in the lysis buffer was used to reduce non-specific DNA binding, and the pH of the buffer was increased to 9 to precipitate some other proteins with lower isoelectric points. Although neither of these measures discernibly improved the yield of *ro*HD as assessed by SDS-PAGE, further fractionation of the soluble extract was attempted.

Ammonium sulphate ($AmS0_4$) fractionation was tried since, being small and basic, the *ro*HD might have been amongst the more soluble proteins. Supernatants were taken to 30 and then 60% $AmS0_4$ saturation, mixed on ice, and centrifuged. Most of the

higher molecular weight protein was removed at 30%, and the *ro*HD was precipitated by 60% $AmSO_4$. However, cation-exchange chromatography on Biorex A-70 (pH 9.0) of the dialysed 60% pellet showed that there to be many low molecular weight proteins (<20 kdal) of fairly equal quantities, of which there were at least three that eluted from the column at >0.5 M NaCl and ran adjacent to the s/n 4+ *ro*HD marker on SDS-PAGE gels. Subsequent chromatography on Mono-S (pH 9.0) of fractions containing these proteins resulted in three half-pure proteins eluting at 0.75, 0.8, and 0.88 M NaCl, all of which were estimated from SDS-PAGE to be present at <10µg/litre of induced culture (not shown).

Attempts to use the X17 polyclonal antisera to localise the distribution of *ro*HD in the various fractions were not especially successful. With Western blotting, there were problems getting satisfactory conditions for transfer of the smaller, basic proteins (see Figure 5.5). Remedies to this problem were tried with limited success (Szewczyk and Kozloff 1985; Section 5.2.6), since the antisera exhibited specificity for other small (<20 kdal) proteins giving difficulty in the interpretation of data. Immunoprecipitation was hampered by similar problems of background. Thus, the only way to assess the presence of the *ro*HD was by visual appraisal of protein bands in SDS-PAGE gels with similar migration to the s/n 4+ marker.

The possibility that the soluble *ro*HD concentration is limited by binding to DNA was investigated by use of a more gentle freeze/thaw/lysozyme lysis procedure, since this does not shear chromosomal DNA as much as does French Press lysis, or sonication. The soluble fraction was further fractionated with Polymin P at a higher concentration than previously used (1%). The resulting supernatant showed no noticeable improvement in yield of *ro*HD; the pellet was extracted with NaCl (2 M) and urea (4 M) to eliminate the possibility that the *ro*HD was being precipitated by

the Polymin P.

Ways of improving the extraction of insoluble *ro*HD from the pellet following lysis were also investigated, with NaCl (2 M), and urea (8 M). However, there was no apparent improvement upon the extraction described in Section 3.2.1. However, these studies indicated that the quantity of *ro*HD in the induced cells prepared in the batch fermenter (40 litre culture) was not as high as that obtained in small (<1 litre) cultures in shaking water baths. It was not clear why this should be so but there were slight variations in the temperature of induction, the smaller cultures being more constant. Temperatures higher than 42°C (upto 45°C) were tested since it had been observed that this gave improved yields of expressed dihydropteridine reductase protein in a pCE30 derivative (Armarego et al.1989), perhaps as a result of the temperature inactivation of proteases. The results of these studies were however inconclusive. The possibility that the *ro*HD may be subject to proteolysis was studied by using an *E. coli* strain with a proteolysis deficient *htpR* mutation (*htpR am165*; Goff et al.1984) as a host strain for pJG(2) and pJG(3). Improvements in the abundance of *ro*HD as detected by SDS-PAGE were again not clear (data not shown).

A main conclusion from this body of work was that, in the absence of an assay for the *ro*HD, it was difficult to distinguish readily between the different treatments, although none of them provided any evidence that there were significantly larger quantities of the protein present in the induced cells than had been indicated by the studies reported in Sections 3.2.1 and 3.2.2. Thus, since there was evidence that more of the *ro*HD was present as insoluble protein which could be extracted with high concentrations of urea, it was decided to attempt to purify the *ro*HD from the lysis pellet by urea extraction.

Figure 3.8: SDS-PAGE and elution profile from BioRex A70 ion-exchange column of the supernatant obtained by urea extraction of the cell lysate pellet.

Preparation of the protein extract: Induced pJG(2) cells (67 g net weight) were suspended in 250 ml lysis buffer (25 mM Tris pH8, 10% sucrose, 2 mM DTT, 0.25 mM PMSF); lysozyme (76 mg) was added, and the mixture stirred at 4°C for 60 minutes, and then incubated at 37°C for 5 minutes. The lysate was centrifuged at 12,000g (60 minutes, 0°C), and the pellet resuspended in 8 M urea buffer (8 M urea, Buffer A, 0.4% Triton X-100) by mixing with a tissue homogeniser and stirring for 3 hours at 4°C. Half of this solution was diluted with 1 volume of 4 M urea in Buffer A (total volume, 220 ml), and then mixed gently with 50 ml BioRex resin (pre-washed with 2 M NaCl, Buffer A; equilibrated with 0.1 M NaCl, Buffer A) for 1 hour at 4°C. This mixture was then poured onto a glass sinter under gentle suction, and washed with 400 ml 0.1 M NaCl/Buffer A. The resin was suspended in 100 ml 0.1 M NaCl/ Buffer A, and poured onto a 10 ml bed of fresh BioRex resin (pre-equilibrated with the buffer) to give a 12 x 2.5 cm column. This column was washed with 50 ml 0.1 M NaCl/Buffer A, and then eluted with 30 ml 1.1 M NaCl/Buffer A, and 100 ml 2 M NaCl/Buffer A. 10 ml elution fractions were collected, their A_{280nm} measured, and protein composition analysed by SDS-PAGE. A preponderance of low M_r proteins were found in the first 4 fractions eluted at 2 M NaCl. These were dialysed against 0.1 M NaCl/Buffer A; this gave a white precipitate (lane **F**) and a supernatant which was loaded onto another BioRex column (lane **E**).

BioRex chromatography: 30 ml BioRex resin was poured into a 6 x 2.5 cm column, pre-equilibrated with 0.1 M NaCl/Buffer A, and then loaded with the protein extract above (= lane **E**). The column was then eluted with: i) 30 ml 4 M urea/Buffer A (eluate = lane **G**); ii) 30 ml 0.1 M NaCl/Buffer A; iii) 50 ml 0.1 M NaCl/10 mM CAPS pH 10.0 (Buffer C) (eluate = lane **H**); iv) a 0.1 to 1.1 M NaCl gradient (100 ml in Buffer C) (1 ml elution fractions = 1 to 70 in the elution profile; lane **I** = fraction 10; **J** = 16; **K** = 21; **L** = 26; **M** = 28; **N** = 32; **O** = 36); v) 100 ml 2 M NaCl/Buffer C (10 ml elution fractions = 71 to 79 in the elution profile; lane **A** = fraction 72; **B** = 73). All loadings on SDS-PAGE (18% gel) are 20 µl samples of fractions + 20 µl SDS buffer. Lane **C** = molecular weight markers; Lane **D** = s/n 4+ (see Figures 3.2 and 3.4). The positions of the *ro*HD in s/n 4+, and HU I/II in lane **C** are indicated.

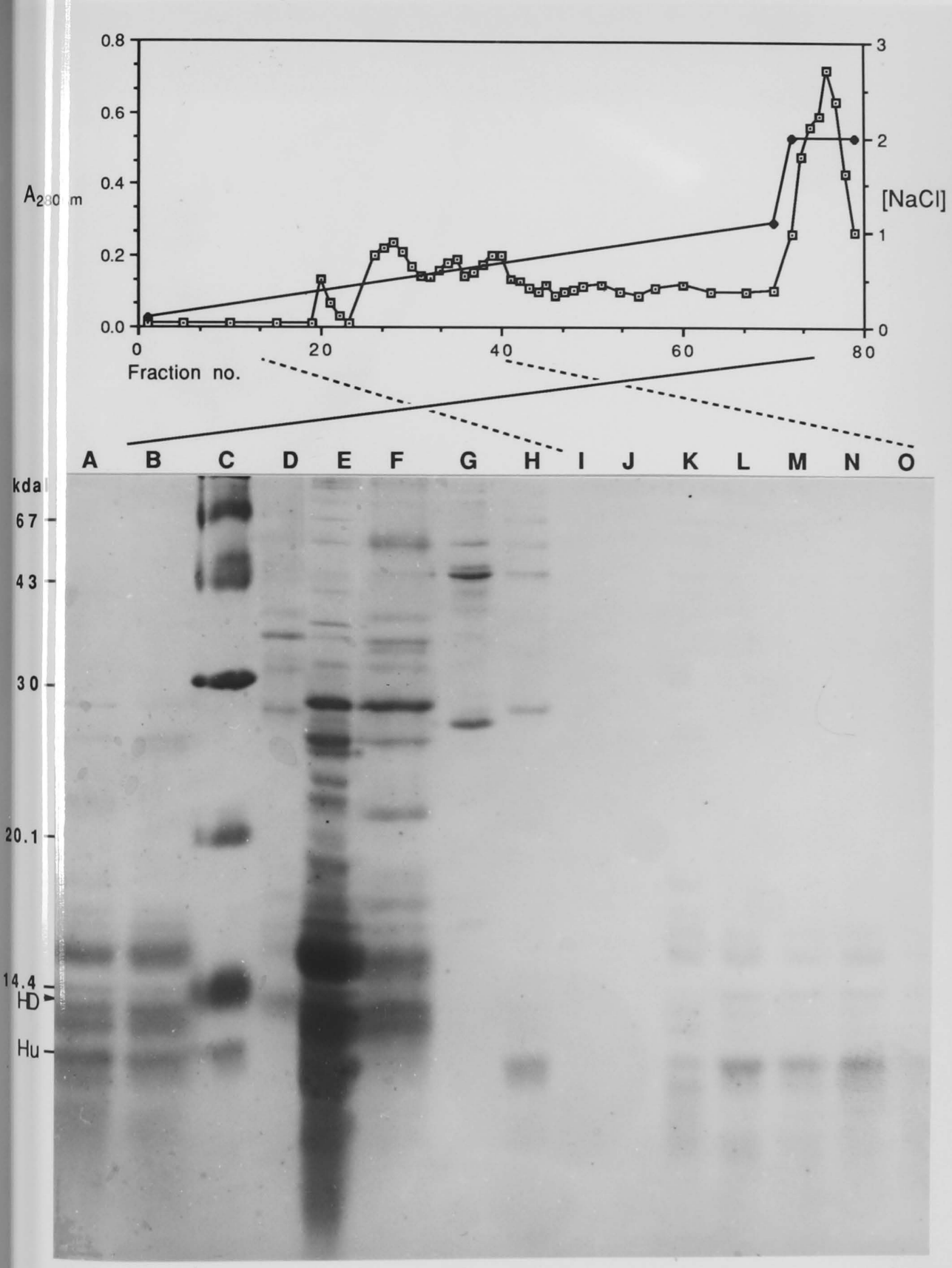


Figure 3.8: SDS-PAGE and elution profile from BioRex A70 ion-exchange column of the supernatant obtained by urea extraction of the cell lysate pellet

3.2.4 Urea Extraction.

The induced AN1459/pJG(2) cells were thawed in phosphate lysis buffer (with no NaCl). Lysozyme was added and the mixture was left stirring at 4°C for an hour after which it had become thick and glutinous. Lysis was effected by placing the mixture at 37°C for 5 minutes. The lysed cells were centrifuged (12,000g, 1 hour, 0°C) give a bright yellow slightly viscous supernatant and a thick, viscous brown pellet. An equal volume of 8 M urea buffer was added to the pellet, and mixed in with a tissue homogeniser, followed by stirring at 4°C for 3 hours until the pellet was fully resuspended. This was then centrifuged to give a thick brown supernatant.

The question now arose as to how to remove the DNA from the extract to facilitate further protein purification. Chromatography was chosen in preference to alternative methods (such as addition of Polymin P, spermidine, DNase, or $AmSO_4$) - it was decided to load the supernatant onto a BioRex column, and then to wash the DNA through, before recovering the protein.

A small sample of the supernatant was loaded onto a small column (2 x 0.7 cm) equilibrated with Buffer A. This was washed extensively with 0.1 M NaCl, and finally eluted with 1.1 M NaCl. SDS-PAGE of the fractions showed that whilst most of the protein in the extract passed through the column, the eluate consisted predominantly of a mixture of around six low molecular weight proteins between 10 and 18 kdal (Figure 3.8). Of these, two comigrated with the *ro*HD in s/n4+, and one with HU I/II.

However, the supernatant was too viscous to permit a ready scaling-up of this chromatography, so it was decided to dilute the extract, before mixing it with the BioRex resin (50 ml) in a beaker. The resin was washed thoroughly with 0.1 M NaCl (Buffer A) on a glass sinter before loading it into a column (12 x 2.5 cm) and

Figure 3.9: SDS-PAGE and elution profile of Mono-S column loaded with fractions from Figure 3.8.

Protein extract: Fractions 71-79 eluted from BioRex column (see Figure 3.8) in 2 M NaCl/Buffer C (lanes A and B in Figure 3.8) were dialysed against 50 mM CAPS pH 10.3 (Buffer D)/0.2 M NaCl (= Load).

Mono-S chromatography: The pre-packed 5 ml Mono-S column was used with the Pharmacia FPLC apparatus. The column was pre-equilibrated with 20 ml 0.2 M NaCl/Buffer D. The above protein extract (20 ml) was applied to the column, which was then washed with 15 ml 0.2 M NaCl/Buffer D (= fractions 1-10 in the elution profile). A steep salt gradient was developed: 5 ml 0.2 to 0.5 M NaCl/Buffer D (= fractions 11-14). This was followed by a flatter gradient: 60 ml 0.5 to 0.9 M NaCl/Buffer D (= fractions 15-54). The A_{280nm} of all 1.5 ml elution fractions was measured, and plotted to give the elution profile. 20 µl samples of these fractions were electrophoresed on 18% SDS-PAGE. The s/n 4+ marker (see Figures 3.2 and 3.4) is shown. The protein in Fraction 31, which has a similar M_r to this marker, was shown by protein sequencing to be the *rough* homeodomain. A sequence could not be obtained for the protein in Fraction 30.

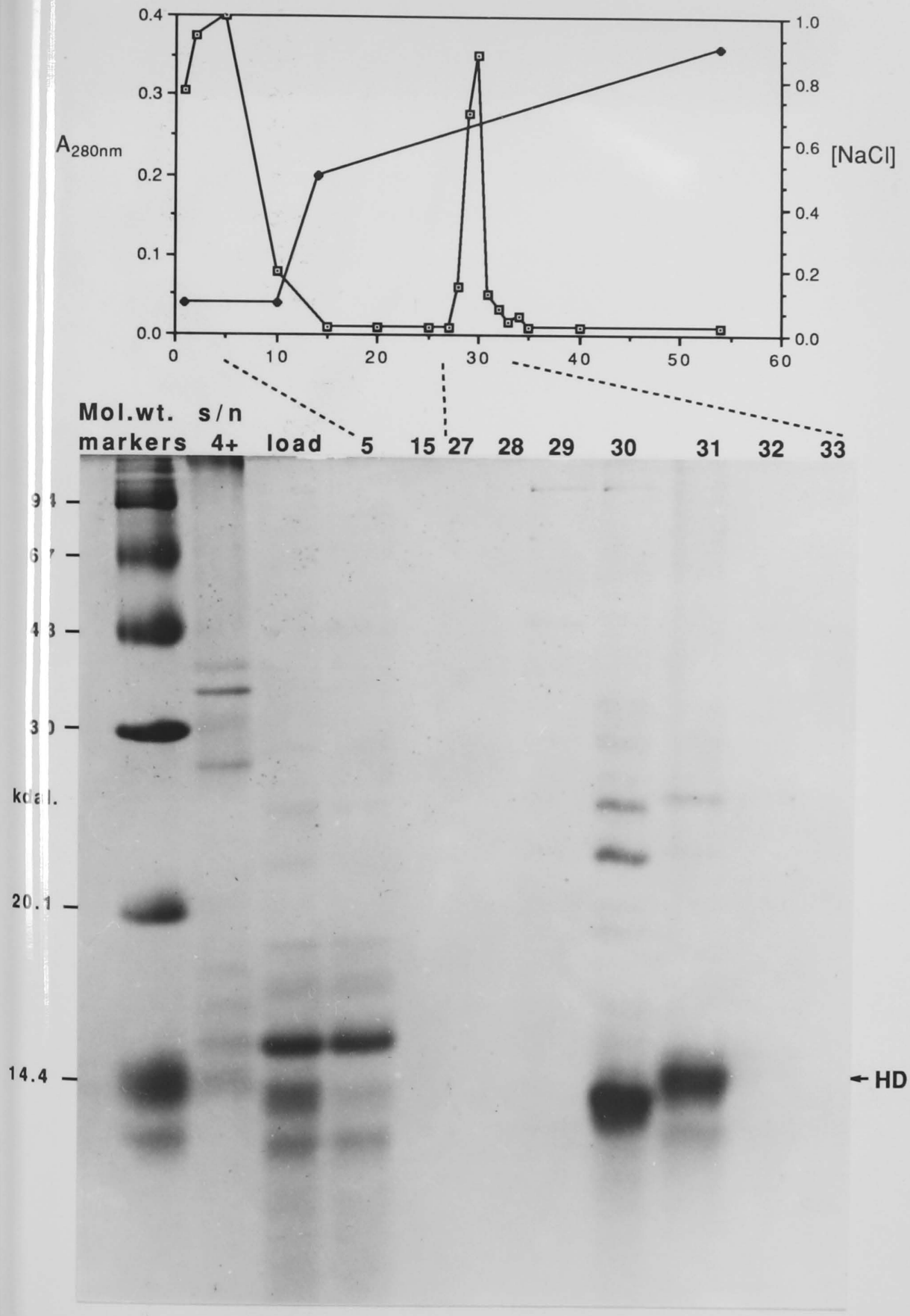


Figure 3.9: SDS-PAGE and elution profile of Mono-S column loaded with fractions shown in Figure 3.8.

eluting bound protein from it with 2 M NaCl. SDS-PAGE of the eluate consisted of a similar mixture of low molecular weight proteins to that eluted from the above column however, there were many other higher molecular weight proteins (> 30 kdal) now evident in the fraction. Some of these additional proteins may still be present due to the relatively faster removal of urea from the resin washed on the glass sinter - perhaps they were renaturing as the urea was removed.

To test this possibilty, the eluate was dialysed to low salt and loaded onto another BioRex column (6 x 2.5 cm; see Figure 3.8), which was washed with 4 M urea solution. The column was re-equilibrated with 0.1 M NaCl, Buffer A before raising the pH with CAPS (Buffer C) to 10.0 (Figure 3.8). This shift to higher pH was designed to remove more proteins by changing their charge, and hence, affinity for the column. At pH 10.0, the column was finally eluted with a 0.1 to 2.0 M NaCl gradient. SDS-PAGE of the various fractions (Figure 3.8) showed that both the passage of urea through the column, and raising the pH of the column, resulted in the elution of the high molecular weight protein. The NaCl elution gradient at pH 10.0 results in a similar mixture of low molecular weight proteins to that obtained with the smaller column (Figure 3.8).

Samples of this latter eluate were applied to a Mono-S column at different pH's (7.3 and 10.0), with a 0.1 M to 1.1 M NaCl gradient. SDS-PAGE of the elution profile showed there to be a fraction at pH 7.3 containing two proteins migrating adjacent to the *ro*HD marker, and some other contaminating proteins; at pH 10.0, the equivalent fraction contained just these two proteins in roughly equal quantities. To separate these two proteins, Mono-S chromatography was repeated at pH 10.3 with a stepped NaCl gradient (0.2 to 0.9 M; see Figure 3.9). The two proteins were successfully separated, eluting at 0.65 M, and 0.66 M NaCl (Figure

3.9). The N-terminal sequence of the first 16 amino acids of the upper protein (0.66 M NaCl; determined by Dr. Shaw), confirmed its identity as the *rough* homeodomain. A sequence was not obtainable for the other protein.

The concentration of the *ro*HD was determined by amino acid composition analysis (Dr. Shaw). This showed there to be a total of 630µg of the homeodomain in the 1.5 ml fraction. This had been obtained from an equivalent of 2 litres of the induced cell culture.

Thus, the urea extraction procedure was successful in purifying the *ro*HD. However, its yield of 315µg per litre of induced culture is not particularly good, especially when compared to the *Antp* HD yield of upto 4 mg per litre (Muller et al.1988). Although considerably larger quantities of *ro*HD were required for the proposed studies of the protein structure, it was considered to be more profitable to postpone further purification of the protein until an investigation of the DNA-binding properties of the *ro*HD. In this way, a specific DNA-binding assay might be developed which could facilitate the necessary bulk purification of the *ro*HD. The study of the *ro*HD's binding to various DNA sequences, and the prospects for an assay are discussed in the next chapter.

3.3 Materials and Methods.

3.3.1 SDS-PAGE.

This involves the denaturing of proteins with the ionic detergent, sodium dodecyl sulphate (SDS), and their subsequent separation by mass/charge ratios on an SDS-denaturing polyacrylamide gel.

Cracking Procedure: The bacterial cell culture was usually spun down and resuspended in SDS-lysis ('cracking') buffer (50mM DTT, 1.2% SDS, 0.5M Tris base, 60% bromophenol blue/glycerol) to an O.D.

of 10, i.e. if the optical density (O.D.) at 595nm (A_{595}) = 1.0, then the cells were resuspended in 1/10th volume crack. This was heated at >90°C for several minutes and 25-40µl loaded per lane of the gel with a Hamilton syringe.

SDS-polyacrylamide gel: The gel consisted of a lower separating gel of an appropriate acrylamide concentration (according to the molecular weight which one wishes to focus on, e.g. 7%=40 to 200 kdal; >15%=5 to 50 kdal), and a higher stacking gel (usually 5% acrylamide) in which the loading wells are formed (Laemmli 1970).

Separating gel: e.g.15%: 15ml 30% acrylamide solution (29:1 acrylamide:bis) + 7.5ml 1.5M Tris HCl pH 8.8 + 0.3ml 10% SDS + 100µl 10% Ammonium persulphate (APS) + 10µl N,N,N',N'-tetramethylethylene diamine (TEMED) + 7.1ml water.

This was poured into a 1.5mm x 200mm x 150mm vertical glass mould and left to set. With a gradient gel, e.g. 7 to 15%, the 7 and 15% solutions were gradually mixed during loading such that the separating gel goes from 7% at the top to 15% at the bottom.

Stacking gel: 0.833ml 30% acrylamide + 1.25ml 0.5M Tris HCl pH6.8 + 50µl 10% SDS + 40µl 10% APS + 4µl TEMED + 2.8ml water. This solution was poured on top of the set separating gel. Combs with an appropriate number of well moulds were placed into the stacking gel mix (12 to 18 wells), and it was left to set.

Molecular weight markers (kdal):

Pharmacia Low Mwt. standards: 94, 67, 43, 30, 20.1, 14.4.

BioRad Pre-Stained standards (kdal): 130, 75, 50, 39, 27, 17.

Gel running: The gel was placed into the electrophoresis apparatus (Berger and Kimmel 1987), and immersed in running buffer (3.6g Tris, 17.3g Glycine, 0.6g SDS, H_2O to 600ml). Once protein samples were loaded, the gel was run at 80 volts until the dye had traversed the stacking gel, and 120 volts for the separating gel; although voltage was varied between 20 and 150 volts to vary the

speed of electrophoresis as desired.

Staining: Gels were stained and fixed by placing them in a Coomassie Blue solution (0.5g Coomassie Blue, 20% methanol, 7.5% acetic acid) for ~1/2 hour. The gels were destained in 10% acetic acid, 10% ethanol until the protein bands were distinct. The stained gel was photographed with a Polaroid land camera using Type 55 film.

3.3.2 Bacterial Cultures.

The plasmid-bearing clones were maintained as DMSO stocks at -70°C (1ml culture + 8% dimethyl sulphoxide in a sterile bottle). They were grown by streaking out on LBAmp agar plates and grown at 37°C (30°C if temperature sensitive, as in the case of pJG(2)).

Induction Procedure: A small (10-50 ml) overnight LBAmp culture was grown at 30°C. This was diluted 20-fold with LBAmp and grown at 30°C until in the exponential growth phase (A_{595} >0.5). The exponential phase culture was induced by transferring it to a bath at 42°C. The culture was then grown at this temperature for varying times before harvesting (usually 2 hours). (One litre induced cultures were usually grown by incubating a 500ml culture at 30°C before adding the remaining 500ml of LBAmp (preheated to 54°C) when the A_{595} >1.0). The cultures were centrifuged at 6000g for 10 minutes at 4°C. The cell pellets were resuspended in the appropriate buffer for lysis.

The 40 litre preparation was carried out in a batch fermenter.

3.3.3 Lysis Procedures.

The cells were lysed to release their contents using one of three methods.

French Press: The cell pellet was resuspended in lysis buffer as indicated. Section 3.2.2 used Buffer A (= 50mM phosphate pH7.5),

10% sucrose, 2mM EDTA, 1mM DTT, 0.4M NaCl. pH 9.0 was attained with Buffer B = 20mM glycine. Varying concentrations of NaCl were used. For French Pressing, the cells were usually resuspended to A_{595} = 50. This suspension was placed in the French Press and crushed at pressures usually in excess of 10,000 psi. The resulting lysate was centrifuged at 12000-16000g.

Lysozyme/Freeze/Thaw: Cells were resuspended in buffer e.g. 25mM Tris pH8, 10% sucrose, 50mM NaCl, 2mM DTT, 0.25mM PMSF. Egg white lysozyme was added to a concentration of 100µg/ml or 1mg/g cells. The solution was mixed at 4°C for >1 hour. In the earlier treatments, the mixture was frozen in liquid nitrogen and then thawed. Subsequently, the mixture was simply warmed to 37°C for a five minutes. The lysates were centrifuged as indicated.

Sonication: The cells were resuspended in buffer (50mM Tris pH7.5, 10% sucrose). A sonication probe was placed in the suspension and sonication was usually carried out at 100 Watts in four 1 minute bursts at 4°C. The lysate was then centrifuged, typically at 12000g.

Post-Lysis Treatments.

Polymin P. This was added dropwise as a 10% solution to give a final concentration of 0.1 to 1% as indicated. The solution was mixed for 15 minutes at 4°C before the white flocculent precipitate formed was spun down by centrifugation.

Ammonium Sulphate This was added to the lysate solutions as a solid to give a final concentration of 10 to 95% of saturation where saturation is 767g per litre.

Urea Extraction of the Pellet. This was with an 8M urea buffer (10mM Tris 7.4; or Buffer A). A tissue homogeniser was used to thoroughly resuspend the pellet.

3.3.4. ^{14}C-Arginine Incorporation.

This entailed growing the bacteria in an appropriately supplemented minimal media in order to maximise the uptake of the added arginine. An AN1459/pJG(2) LBAmp culture at 30°C was grown until A_{595} >0.4. This was centrifuged and the cells were resuspended in an equal volume of supplemented minimal media (= 5ml 20x56 solution (40g/l $(NH_4)_2SO_4$; 212g/l K_2HPO_4; 122g/l NaH_2PO_4, pH 7.1) + 0.1ml 1000 x's Trace salts (0.01mM $ZnSO_4$, 0.001mM $(NH_4)_6Mo_7O_{24}$, 0.4mM H_3BO_3, 0.03mM $CoSO_4$, 0.01mM $CuSO_4$, 0.08mM $MnSO_4$, 2mM $FeCl_2$, 2.5mM $CaCl_2$, 10mM citric acid) + 0.1ml Vitamin B + 0.1ml 1M $MgCl_2$ + 2ml 2M glucose + 1M 1% Theonine/0.5% leucine + 1ml 0.5% Isoleucine/0.5% valine. Water to 100ml. Ampicillin was added (25µg/ml). This was grown at 30°C for 15 minutes, then 1ml aliquots were induced by growing at 42°C. 50µl [^{14}C]-Arginine solution was added at varying times after induction (50µCi/ml from Amersham). The cells were pelleted 30 minutes after addition of arginine, resuspended in SDS-lysis buffer at A_{595} =10.0, and separated by electrophoresis on 20% SDS-PAGE gels until the dye reached the bottom. The gel was then dried under vacuum and autoradiographed for 2 days.

3.3.5 Western Blotting/Immunoprecipitation.

These methods are discussed in Chapter 5.

3.3.6 PVDF Blot/Protein Sequencing.

PVDF blot (Matsudiara 1987)-This was essentially a Western blot onto polyvinylidene difluoride (PVDF) rather than nitrocellulose membrane. AnSDS-PAGE gel (10 to 13%) was run with the protein extracts and molecular weight standards. The PVDF membrane (Immobilon from Millipore) was pre-wetted in methanol, rinsed in water and finally equilibrated in transfer buffer. A blotting

sandwich was assembled with, from the anode end: scotchbrite, 3MM paper, the gel, two sheets of PVDF, 3MM paper, scotchbrite. This was loaded into the BioRad Transblot apparatus with the blotting buffer (10mM 3-(cyclohexylamino)-1-propane sulphonic acid (CAPS) pH11.0, 20% methanol). The blot was at 0.2Amps for 1 hour at 4°C . The membrane was stained in Coomassie blue for 5 minutes. It was then destained in 40% methanol, 10% acetic acid overnight, after which it was air-dried.

Protein Sequencing: This was done by Dr Dennis Shaw, JCSMR, A.N.U. using an Applied Biosystems Model 477 Pulsed Liquid Phase Protein Sequencer with on-line 120 PTH Analyser.

The N-terminal sequence of the *ro*HD in Section 3.3.4 was obtained with 50 pmol of the protein, with a repetitive yield of 95% per residue.

3.3.7 Ion-Exchange Chromatography.

Two cation-exchange resins were used in these experiments: BioRex A-70 (BioRad), and Mono-S (Pharmacia). The BioRex A-70 resin was soaked and packed into columns with a bed volume from 2 to 60ml. Mono-S came as a pre-packed 5ml column for use on the Pharmacia Fast Protein Liquid Chromatography (FPLC) system. Thus, whilst the BioRex columns were run with peristaltic pumps, and salt gradients were mostly set up using salt bridges between a low and high NaCl solution, Mono-S columns were run with computer-controlled pumps and NaCl gradients. Typically, the column resins were pre-equilibrated prior to loading by washing with 2 volumes of high salt buffer and at least 2 volumes of low salt buffer. Samples were loaded directly onto the columns. Once the columns were loaded, they were extensively washed with low salt buffer until the solution coming off the column had returned to the A_{280} prior to loading as shown by a U.V. path monitor and a

chart recorder. The appropriate NaCl gradient was now run, usually from a linear low salt (0.1M NaCl) to high salt (1.1M) buffer. Buffers used were Buffer A = 50mM phosphate pH7.4 (NaH_2PO_4/Na_2HPO_4); Buffer B = 20mM glycine pH9 (glycine, NaOH); Buffer C = 10mM CAPS pH 10.0 (CAPS, NaOH); or Buffer D = 50mM CAPS pH 10.3 (CAPS, NaOH). The eluates were collected in a fraction collector as 1ml fractions for small columns, upto 10ml for large columns.

The elution profile shown by the U.V. path monitor was checked by obtaining the A_{280} of various samples in a UV spectrophotometer.

Dialysis: All dialysis steps used Spectra/Por molecular dialysis membrane tubing 3 (Spectrum Medical Industries) with a molecular weight cut-off of 3,500. Dialysis was typically in 2 x 1 litre of the stated buffer.

CHAPTER 4.

DNA-Binding Studies.

4.1 Introduction

4.1.1 Principles of DNA-Protein Interaction

The determinants of DNA-protein interaction relate to a number of factors including the DNA and protein structure, their relative concentrations, and the ionic environment, the process being comparable to the binding of any ligand (whether an enzymatic substrate or allosteric effector) to a protein molecule.

Two aspects of the DNA structure are crucial determinants of its specific interaction with DNA-binding proteins. First, the sequence-dependent polymorphism of DNA allows the protein to recognise structural features, both global and local, of the double helix in addition to direct interactions with individual base pairs. The relative contributions of these two components is highly variable, ranging from such proteins as those in the core nucleosome particle for which structural recognition is dominant, to the λ repressor where selectivity is determined principally by base-specific interactions (Drew and McCall 1988; Travers and Klug 1987). In between, there are proteins such as catabolite activator protein (CAP) and the phage 434 repressor, for which both factors are significant (Travers 1989).

The other major property of DNA is its conformational flexibility. This is utilized in interactions with proteins both to promote enzymatic manipulation and also to allow a snug fit between the protein surface and the double helix. Here, the DNA is not always passive, for example, it appears that in the case of the histone octamer and the *trp* repressor, the sequence of the DNA binding site can also influence the conformation of the bound

protein (Travers 1989).

The variations in the DNA then are primarily related to the sequence of the nitrogenous bases, although there is evidence for non-sequence-specific protein-DNA interactions (e.g. with histones) (Drew and Travers 1985). The selective binding of a protein to a particular DNA sequence requires the recognition by the protein of an ensemble of steric and chemical features that in total delineate the binding site.

Most of the DNA in cells is thought to exist in the B form, in which the helix is twisted asymmetrically to give a smaller minor groove and a larger major groove (Drew and McCall 1988). In principle, the DNA sequence can be "read" directly by hydrogen-bonding from the major groove of B form DNA without disruption of the double-stranded structure. The hydrogen-bond donors and hydrogen-bond acceptors of the bases are exposed in both the major and minor grooves, although from a steric viewpoint, the major groove is favoured for interaction. With a reasonable degree of flexibility in the protein structure, and utilizing only hydrogen bonding to the bases, the formation of a minimum of two hydrogen bonds in the major groove is necessary to unambiguously read the identity of a base pair (Seeman et al.1976).

Thus, it has been noted that the detection of a hydrogen-bond acceptor at the acceptor position on a thymine (T) does not distinguish the T from a G (guanine). In the minor groove, A (adenine) and T are indistinguishable and G and C (cytosine) are barely distinguishable; however, an A-T (or T-A) base pair can be distinguished from a G-C (or C-G) base pair. The methyl group on thymine provides an additional sequence recognition element, that can be identified from the major groove.

Since the major groove of B form DNA will nicely accommodate an α-helix of a protein (an α-helix is about 12 Å in diameter, and the major groove is 6-8 Å deep, and about 12 Å

wide), then it is proposed that suitable hydrogen bond-forming amino acid side-chains on an exterior face of the helix should be capable of "reading" a DNA sequence (Schlief 1988).

Sequence information can also be expressed by means of structural inhomogeneties along the DNA helix due to the base sequence. Such sequence-dependent structural variations in the repeating units along the DNA helix have been shown by X-ray crystallography of synthetic oligonucleotides (Drew 1988). Each of the repeating units of phosphate-sugar-base along one strand of DNA possesses six bonds about which rotation is possible, in addition to the various pucker conformations of the sugar ring. The helical twist from one nucleotide to the next, the tilt of a base pair around an axis within the plane of the hydrogen bonds between the bases, the propeller twist of one base with respect to the other in a base pair, and the buckle of the bases out of one plane after removal of the propeller pitch, all vary from one base pair to the next along the DNA.

Thus, it ought to be possible for a protein "reading" the DNA sequence to ignore the chemical or hydrogen bonding differences between the four bases along the DNA (see above), and instead to read information directly or to bind to a specific site by recognition of the overall structure generated by a sequence, e.g. by detecting the precise positions of phosphates along the DNA backbone (Schlief 1988).

As discussed earlier, proteins seem to show a wide variability in their interactions with DNA, and so may read the DNA sequence by utilizing both the pattern of base-specific hydrogen bonding and methyl groups, and by recognising the overall structure.

A number of general points have been made regarding the structure of proteins capable of sequence-specific binding to DNA. Such proteins usually have a regulatory function, and hence must

possess great selectivity for the correct binding site, binding with high affinity to the correct site and with relatively low affinity to most other sites. It has been suggested that this binding selectivity can be increased for a monomeric DNA-binding protein by the formation of multimers (Schlief 1988). This is explained in terms of the amount of binding energy a protein generates when it makes a fixed number of hydrogen bonds and van der Waals interactions with the DNA. In the process of binding, the protein's centre of mass must be located properly with respect to the DNA site and the protein must assume the correct angular orientation. It is also likely that the protein and DNA undergo conformational fluctuations, the assumption being made that only a subset of the protein's conformational states is capable of binding to the DNA site. This conformational flexibility serves to reduce the effective concentration of the protein on the one hand, and requires binding energy to drive the protein into the correct conformational state on the other. Thus, with a fixed set of interactions, binding can be increased by any means that holds the protein in the correct orientation, or that maintains the protein in the correct conformational state.

It has been found that many protein-binding sites on DNA are symmetric, and that the proteins which bind to these sites possess either two or four identical subunits (e.g. Cro, *lac* repressor; Pabo and Sauer 1984). A dimeric protein binds more tightly than is indicated by the affinity of the individual monomers. For a typical monomeric protein, there is a substantial decrease in entropy required for it to assume the correct position and orientation for binding. However, similar entropic requirements apply in positioning and orienting a dimer. Thus, for a dimer, the binding of one monomer correctly positions the second monomer so that its binding interactions can be used primarily to increase the tightness of binding rather than in properly locating

or orienting the protein. The extra affinity provided by two connected binding domains has been referred to as cooperativity. The interaction energy between monomers generates cooperativity in binding so that the concentration of monomer required to achieve binding to both sites is lowered. Thus, if a protein which binds tightly is one that is more often bound than unbound, a dimer binds tightly because when one monomer momentarily dissociates from the DNA, the other monomer is still likely to remain associated (Ptashne et al.1982).

Another consideration is the displacement of ions by DNA-binding proteins (Schlief 1988). This arises because of the electrostatic attraction between the anionic phosphate groups of the DNA backbone and cations or molecules in solution (or in the cell). This serves to increase the local concentration of the charge-neutralizing cations along the DNA backbone. Since DNA-binding proteins are often themselves positively charged in their binding regions, this will tend to interfere with the binding process. If, as a protein binds to a specific site on the DNA, it displaces the neutralizing charges, then they can be considered to be part of the binding reaction. The equilibrium binding constant depends not only on the concentration of the protein and DNA, but also on the concentration of the charge-neutralizing ion.

Since as many as 10 to 15 such ions may be displaced as a protein binds to DNA, the affinity of a protein for DNA can vary markedly with ionic strength. For example, the affinity of *lac* repressor for the *lac* operator DNA has been found to reduce *in vitro* by a factor of 20 with an increase in NaCl concentration from 0.1 to 0.2 M (Schief 1988). However, *in vivo*, it has been suggested that there may be compensating factors. For example, the main ions near the phosphate of DNA could be polycations whose concentration is not so readily varied; and many proteins may bind ions to themselves during the binding process to compensate for

those displaced from the DNA (Schlief 1988).

Finally, it appears that sequence-specific DNA-binding proteins may possess one of several highly-conserved types of binding domain with which they recognise and interact with DNA; for example, "zinc fingers", the "helix-turn-helix" motif, "leucine zippers" and "POU" domains (Schlief 1988; Pabo and Sauer,1984; O'Shea et al.1989; He et al.1989).

The first of these to be structurally characterized, the "helix-turn-helix" motif, consists of a number of α-helices, notably two short α-helices with a short, sharp β-turn, one of which lies in the major groove of the DNA, the other lying across it and also apparently interacting with the DNA. This motif, which is proposed for the homeodomain, is discussed in the next section.

4.1.2 The α-Helix-β-turn-α-helix Motif.

The first site-specific DNA-binding proteins whose structures were determined by X-ray crystallography were the CAP protein of *E. coli*, lamdba Cro, and the N-terminal domain of lambda repressor (Pabo and Sauer 1984). These prokaryotic proteins bind as dimers to DNA sequences about 15 to 20 base pairs long, and the binding sites often have dyad symmetry (see Figure 4.1a)). Although these were crystallised in the absence of DNA, the DNA-contacting portion of the dimeric protein was ascribed to surface protusions separated by 35 A (the same centre-to-centre distance that separates successive major grooves of B DNA) and orientated at approximately the same angle as the major groove in DNA. This protusion consists of a short α-helix of 8 amino acids (which could be oriented across the major groove), a three-residue turn, and a second helix of 9 amino acids that was predicted to lie partly within the major groove where it could make specific residue-base interactions. This second helix was termed the

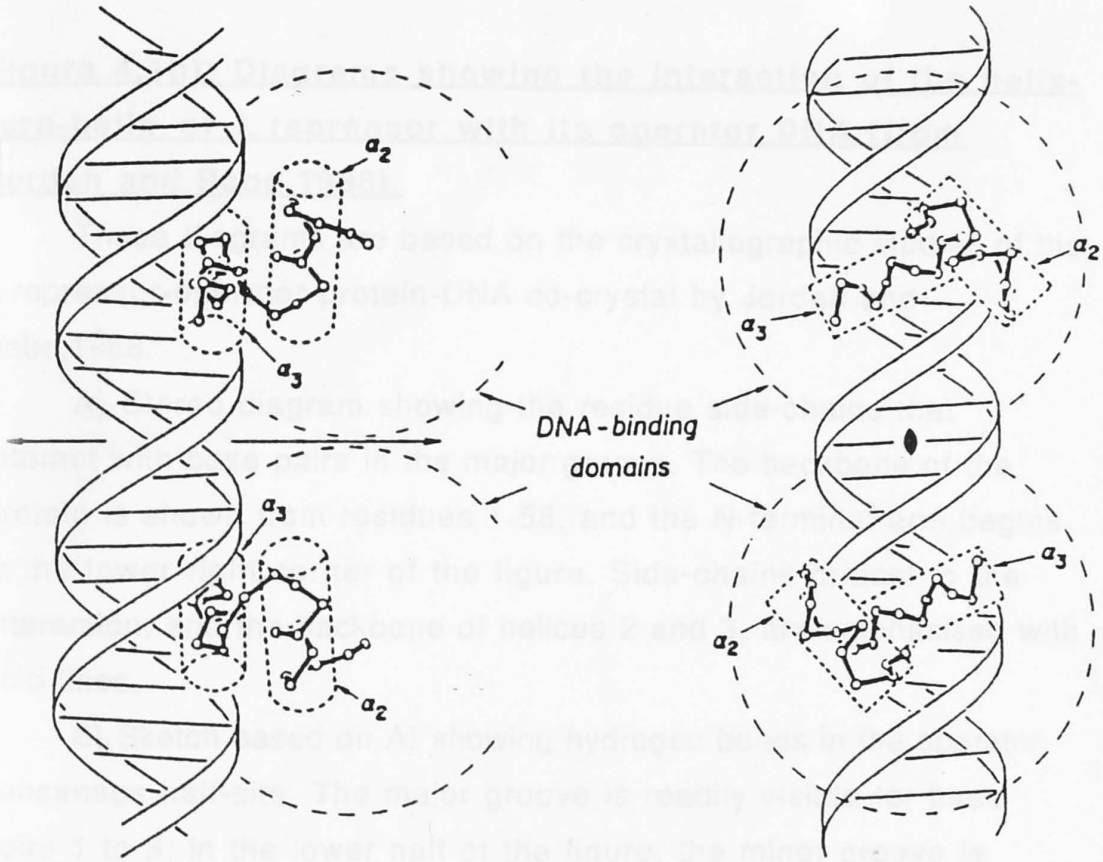


Figure 4.1a): Illustration of the interaction of the helix-turn-helix motif of DNA-regulatory proteins with right-handed B-form DNA (from Ohlendorf and Matthews,1983).

The figure illustrates the general nature of the interaction presumed to occur in many DNA-regulatory proteins between a common α_2-α_3 helical unit and right-handed B-form DNA. The α-carbons from helices two and three of the λ Cro-operator complex are shown. The λ Cro protein binds the operator site as a dimer. There have been a number of crystallographic studies of prokaryotic protein-DNA complexes which clarify this structural interaction (Jordan and Pabo,1988;Aggarwal et al.1988)

Figure 4.1b): Diagrams showing the interaction of the helix-turn-helix of λ repressor with its operator DNA (from Jordan and Pabo,1988).

These diagrams are based on the crystallographic studies of the λ repressor-operator protein-DNA co-crystal by Jordan and Pabo,1988.

A) Stereo diagram showing the residue side-chains that interact with base pairs in the major groove. The backbone of the protein is shown from residues 1-58, and the N-terminal arm begins in the lower right corner of the figure. Side-chains critical to the interaction, and the backbone of helices 2 and 3, are emphasised with bold lines.

B) Sketch based on A) showing hydrogen bonds in the operator consensus half-site. The major groove is readily visible for base pairs 1 to 3; in the lower half of the figure, the minor groove is nearest to the viewer.

C) Sketches showing the hydrogen bonds between side-chains and base pairs 2, 4, and 6 in B)

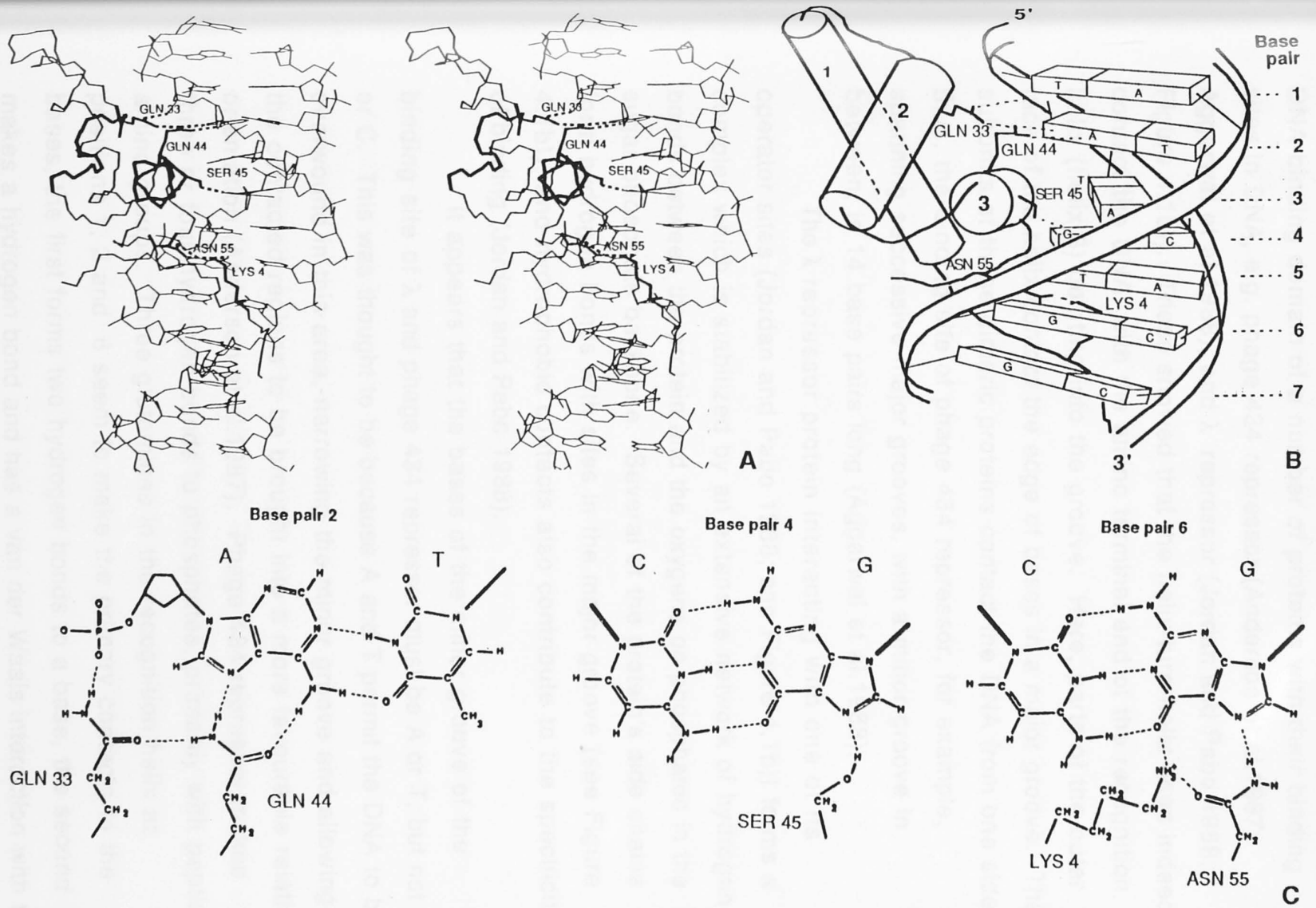


Figure 4.1b): Diagrams showing the interaction of the helix-turn-helix of λ repressor with its operator DNA (from Jordan and Pabo,1988).

recognition helix (Figure 4.1).

Structures have been derived from co-crystals of the DNA-binding domain of a number of proteins with their binding sites in DNA, e.g. phage 434 repressor (Anderson et al.1987; Aggarwal et al.1988), and λ repressor (Jordan and Pabo 1988; Figure 4.1b)). These showed that the helix-turn-helix does indeed contact the DNA, with the amino terminal end of the recognition helix (helix 3) pointing into the groove. Here, parts of the outer face of the helix contact the edge of bases in a major groove. The subunits of these dimeric proteins contact the DNA from one side; thus, the binding site of phage 434 repressor, for example, spanning successive major grooves, with a minor groove in between, is 14 base pairs long (Aggarwal et al.1988).

The λ repressor protein interacting with one of its operator sites (Jordan and Pabo 1988; see Figure 4.1b)) forms a complex which is stabilized by an extensive network of hydrogen bonds between the protein and the oxygens on phosphates in the sugar-phosphate backbone. Several of the protein's side chains form hydrogen bonds with sites in the major groove (see Figure 4.1b)), and hydrophobic contacts also contribute to the specificity of binding (Jordan and Pabo 1988).

It appears that the bases of the minor groove of the binding site of λ and phage 434 repressor must be A or T, but not G or C. This was thought to be because A and T permit the DNA to be overwound in this area, narrowing the minor groove and allowing the contacted regions to be brought into a more favourable relative orientation (Anderson et al.1987). Phage 434 repressor makes three or four hydrogen bonds to phosphates, primarily with peptide amino groups. Three glutamines in the recognition helix at positions 1, 2 and 6 seem to make the primary contacts to the bases, the first forms two hydrogen bonds to a base, the second makes a hydrogen bond and has a van der Waals interaction with the

methyl group on a thymine, and the last makes a single H-bond to each of two consecutive bases (Anderson et al.1987; Aggarwal et al.1988).

The structures of CAP, Cro, λ repressor and 434 repressor in solution display significant differences in the orientation of their recognition helices relative to the major groove (Schlief 1988). Thus, the structure of Cro is most compatible with the recognition helix being parallel and tangential to the major groove of DNA whilst, in λ repressor, the amino terminal end of the recognition helix tends to point inward toward the DNA and thus, its more important DNA contacts derive from amino acids near this end of the helix (Schlief 1988; Jordan and Pabo 1988).

It was noted that the amino acid sequences of the helix-turn-helix regions of Cro, λ repressor, and CAP are similar to each other, and to a number of other bacterial and yeast DNA-binding proteins (Pabo and Sauer 1984). It was such sequence homology which first gave an indication of the likely structure and function of the homeodomain (Laughon and Scott 1984).

4.1.3 Homeodomains as DNA-binding Motifs.

The amino acid sequence towards the N-termini of the proposed homeodomains for *fushi tarazu* (*ftz*), *Antennapedia* (*Antp*), and *Ultrabithorax* (*Ubx*) was found to resemble the conserved pattern of amino acids in the helix-turn-helix region of the prokaryotic DNA-binding proteins (Laughon et al.1984). It was also noted that, given that 30% of the residues in the homeodomain are basic, it is highly likely to associate with nucleic acids.

The amino acids considered to be important for the conformation of the putative helix-turn-helix structure, and hence, those which are conserved, were identified as follows (Laughon and Scott 1984): in the first 8 amino acid helix (helix 2 in the

Figure 4.2: The helix-turn-helix regions in several homeodomains and prokaryotic DNA-binding proteins.

The helix-turn-helix motif within the proteins listed is given in the one-letter amino acid code. Dashes show residues which are identical to *Antennapedia*. The sequences of the homeodomains are from Scott et al.(1989) (*rough* to *cut*, all of these except OCT-2 are *Drosophila* proteins); those for the prokaryotic DNA-binding proteins (λ repressor to CAP), and the yeast MAT α-1 protein are from Laughon and Scott (1984). Helix 2 in the prokaryotic DNA-binding proteins and the yeast MAT α-1 protein is 8 amino acids long; helix 3 is 9 amino acids in length. In the homeodomain, both of these helices are longer, the structure of the *Antennapedia* homeodomain (Qian et al.1989; see Chapter 6) shows them to be 11 amino acids in length, and this is the numbering used.

There have been a number of classifications of the homeodomains into classes on the basis of their amino acid homology. Scott et al.(1989) consider the entire 61 residues of the homeodomain to produce 11 classes: ANTP (including *Antp*, *Ubx*, and *ftz*); DFD (includes *Dfd*); *labial*; *Abd B*; *en* (includes *engrailed*); *eve* (includes *even-skipped*); *prd* (includes *paired*); *hox 1.5*; *hox 2.4*;POU (includes OCT-2); and unclassified (including *ro*, *bcd*, and *cut*). Evidence that a large degree of homeodomain DNA-binding specificity is conferred by residue 9 of the recognition helix has led Hanes and Brent (1989) to propose 5 classes: *Antp*, with Gln at residue 9 (including *Antp*, *Ubx*, *ftz*, *en*, *eve*, *Dfd*, and *ro*); *bcd*, with Lys; *prd*, with Ser; *cut*, with His; and POU, with Cys.

Figure 4.2	Helix 2											Turn			Helix 3 (recognition helix)										
Protein	1	2	3	4	5	6	7	8	9	10	11	1	2	3	1	2	3	4	5	6	7	8	9	10	11

Figure 4.2

Protein	Helix 2											Turn			Helix 3 (recognition helix)										
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	1	2	3	4	5	6	7	8	9	10	11
rough	R	S	R	R	F	E	L	A	E	T	L	R	L	T	E	T	Q	I	K	I	W	F	Q	N	R
Antennapedia	R	R	R	R	I	E	I	A	H	A	L	C	L	T	E	R	Q	I	K	I	W	F	Q	N	R
Ultrabithorax	-	-	-	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
fushi tarazu	-	-	-	-	-	D	-	-	N	-	-	S	-	S	-	-	-	-	-	-	-	-	-	-	-
engrailed	E	-	-	-	Q	Q	L	S	S	E	-	G	-	N	-	A	-	-	-	-	-	-	-	-	K
even-skipped	-	P	-	-	C	-	L	-	A	Q	-	N	-	P	-	S	T	-	-	V	-	-	-	-	-
OCT-2 (POU)	S	E	E	I	L	L	-	-	E	Q	-	H	M	E	K	E	V	V	R	V	-	-	C	-	-
bicoid	A	P	-	L	A	D	L	S	A	K	-	A	-	G	T	A	-	V	-	-	-	-	K	-	-
Deformed	-	-	-	-	-	-	-	-	-	T	-	V	-	S	-	-	-	-	-	-	-	-	-	-	-
paired	I	Y	T	-	E	-	L	-	Q	R	T	N	-	-	-	A	R	-	Q	V	-	-	S	-	-
cut	V	G	T	I	E	F	L	-	N	E	-	G	-	A	T	-	T	-	T	N	-	-	H	-	H
λ repressor				Q	E	S	V	A	P	K	M	G	M	G	Q	S	G	V	G	A	L	F	N		
λ Cro				Q	T	K	T	A	K	P	L	G	V	Y	Q	S	A	I	N	K	A	I	H		
434 repressor				Q	A	E	L	A	Q	K	V	G	T	T	Q	Q	S	I	E	Q	L	E	N		
CAP				R	Q	E	I	G	Q	I	V	G	C	S	R	E	T	V	G	R	I	L	K		
MAT α-1				K	E	E	V	A	K	K	C	G	I	T	P	L	Q	V	R	V	W	F	I		
				1	2	3	4	5	6	7	8	1	2	3	1	2	3	4	5	6	7	8	9		

prokaryotic proteins), there is an alanine at position 5, and a hydrophobic residue at positions 4 and 8; for the correct formation of the β-turn, the side chains need to be small in the last residue of helix 2, and the first two of the turn, the latter of these residues being hydrophobic; in helix 3 (the 9 amino acid "recognition helix"), there is an isoleucine or a valine at position 4, with a hydrophobic residue at position 7. It is thought that the last residue of the turn and residues 1, 2, 3, 5, 6, 8 and 9 of the recognition helix are those whose variability serves to determine the specificity of DNA binding (Figure 4.2).

Thus, it was considered that the homeodomain bore the helix-turn-helix motif and hence, it was presumed that it would act as a sequence-specific DNA-binding domain in the proteins in which it occurred. It was also predicted that, since the specificity of DNA binding was determined by the side groups protuding from the residues of the recognition helix into the major groove, and that groups were identical in this region for *ftz*, *Antp*, and *Ubx* homeodomains, all three should recognise and bind to the same DNA sequence.

The first functional evidence for homeodomains as sequence-specific DNA-binding regions came with the use of β-galactosidase fusion proteins bearing the homeodomain from *engrailed* (*en*) (Desplan et al.1985). Only those fusion proteins with the homeodomain, as against other regions of the *en* protein, displayed sequence-specific binding, although the actual sequence recognised was not characterised. Subsequent DNase I footprinting studies with fusion protein/homeodomain hydrids and, more recently, with the entire protein encoded by cDNAs, have revealed that there are a number of DNA sequences to which the homeodomain will bind *in vitro*, all of which have been found in the promoter regions of genes. These sequences are summarised in Table 4.1 (Scott et al.1989).

Consensus sequence	Specifically bound by
TCAATTAAAT (NP)	***eve,prd,en,zen*** (Hoey and Levine,1988) ***Abd-B, Ubx,*** **OCT-2** (Thali et al.1988) ***en, ftz*** (Desplan et al.1988) ***XlHbox 1*** (Cho et al.1988)
$(TAA)_n$	***ftz, en*** (Desplan et al.1988) ***Ubx*** (Beachy et al.1988) ***Antp*** (Mihara and Kaiser,1988)
TTTGACT	***prd*** (Treisman et al.1989)
TCTAATCCC	***bcd*** (Driever and Nusslein-Volhard,1989)
TTTATG	***cad*** (Dearolf et al.1989)
ANNNNCATTA	***Antp, ftz*** (Muller at al.1988)
ATGCAAATNA/ TNATTTGCAT	**OCT-1, OCT-2** (Muller,M.M. et al.1988) ***Abd B, Ubx*** (Thali et al.1988)
$^{T}_{A}{}^{T}_{A}$TATNCAT	***Pit-1*** (Bodner,1988)
$(TAATCG)_n$	***Ubx*** (Beachy et al.1988)
TCAGCACCG	***eve, prd*** (Hoey and Levine,1988)
AAGGGGTTAA	***Kruppel*** (Treisman and Desplan,1989)
ACNCAAAAAANTA	***Hunchback*** (Treisman and Desplan,1989)

Table 4.1: Homeodomain protein consensus sequences.

The structure determined by 2-D N.M.R. analysis of the *Antp* homeodomain has indicated that it does indeed possess the helix-turn-helix motif, together with a number of other features that may be relevant to its DNA binding ability (Qian et al.1989, see also Chapter 6).

The nature of the specificity of the homeodomain has been examined by means of amino acid substitutions (Hanes and Brent 1989), and by homeodomain substitution (Kuziora and McGinnis 1989).

It is not known to what degree the homeodomain, as opposed to the other regions of homeodomain proteins, confers the target specificity observed during *Drosophila* developmental gene regulation (Kuziora and McGinnis 1989). To test the role of the homeodomain in determining target specificity, Kuziora and McGinnis (1989), replaced the homeobox of the *Deformed* (*Dfd*) gene's cDNA with the homeobox from the *Ultrabithorax* (*Ubx*) gene. The resulting chimeric protein was unable to activate transcription *in vivo* from the *Dfd* gene in *Drosophila* embryos (as detected immunologically), unlike the normal *Dfd* protein, which autoactivates its transcription unit. Instead, the chimeric protein activates ectopic transcription of *Antennapedia*, a gene normally regulated by *Ubx*, and not *Dfd*. This experiment indicates that two closely related homeodomains can confer two distinct regulatory specificities on the same host protein when assayed in the embryo. Thus, similar homeodomains (Figure 4.2) can display a significant amount of target specificity *in vivo*.

The studies of the homeodomain recognition helix by amino acid substitution in the *bicoid* (*bcd*) HD (Hanes and Brent 1989; Treisman, 1989) parallel those in which the binding specificity of the recognition helix of the phage 434 repressor was changed to that of the P22 repressor (both *in vivo* and *in vitro*) by replacement of the external (solvent-exposed) residues of the 434

helix at positions -1, 1, 2, 5, 6 and 9 with those found in P22 (Wharton et al.1985). Similarly, with *bcd*, substitutions were made at positions 1, 2, 5, 6 and 9 of the proposed recognition helix sequence. Firstly, residues 1(Thr), 5 (Lys), 6 (Ile) and 9 (Lys) were all changed to alanines, this being a small residue which does not hydrogen bond to the DNA and has been shown not to disrupt other helices - only with the Lys to Ala substitution at position 9 was the ability to bind to the *bcd* site lost. Secondly, these residues were changed to those found in the *Ubx* recognition helix (thus, at 1, Thr was replaced by Glu; 2, Ala by Arg; and 9, Lys by Gln; residues 3, 5, 6, 7, 8 and 10 are the same in both the *bcd* and *Ubx* helices, and 4 is a conserved Val to Ile substitution). It was discovered that of these changes, only that at position 9 resulted in a change in binding specificity. The modified homeodomain no longer bound the *bcd* site, TCTAATCCC, but instead recognised the *en* site, TCAATTAAAT, which is bound by the *Antp* class of homeodomains (Table 4.1). Based on this observation, it was noted that there are, to date, 5 classes of homeodomains with different residues at position 9: the *Antp* (Gln), the *bcd* (Lys), the *paired* (Ser), the *cut* (His), and the POU (Cys) homeodomain classes (See Figure 4.2).

Treisman et al (1989) made a series of mutations in the helix-turn-helix motif of the *paired* (*prd*) protein. *prd* does not recognise sequences bound by *ftz* or *bcd* (Table 4.1). However, they also found that by changing a single amino acid at position 9 of the recognition helix, the DNA binding specificity was changed. Thus, of all the amino acid substitutions generated in the *prd* helix-turn-helix HD region, only the changes at position 9: Ser (*prd*) to Glu (*ftz*), and Ser (*prd*) to Lys (*bcd*), resulted in significant changes in the binding specificity of the mutant protein. The former mutant bound to the NP_6 sequence (Table 4.1) as tightly as *ftz* , and the latter bound to the *hunchback* promoter as tightly as *bcd*.

Changing all of the helix-turn-helix except position 9 of the recognition helix from *prd* to *ftz* only gave a much weaker (< 20 times) affinity for the NP_6 sequence. However, this may show that the other amino acids of the helix-turn-helix also exert some weak influence on the specificity of the HD (Treisman et al.1989). Thus, there is further evidence for the division of the homeodomains into classes on the basis of the amino acid at position 9 of the recognition helix

As can be seen in Figure 4.2, the *rough* homeodomain has a recognition helix which closely resembles that of the *Antp* class. Hence one might expect the *ro*HD to display a similar binding specificity, even though over the entire homeodomain, they are identical at only 56% of residues (66% if conserved substitutions are allowed) (Scott et al.1989).

However, whilst the five classes are proposed on the basis of residue 9 of the recognition helix, the POU proteins have already been shown to be capable of binding to the *en* (*Antp* class) binding site, TCAATTAAAT (Scheidereit et al.1988; Table 4.1), and the *Ubx* and *Abdominal-B* proteins (with an *Antp*-like Gln at residue 9) will bind to the mammalian octamer site, ATTTGCAT (Thali et al.1988).

Yet, despite such studies *in vitro* which tend to indicate that similar homeodomains are likely to have similar target specificities, the results of the homeodomain substitution experiments in the *Drosophila* embryo described above (Kuziora and McGinnis 1989) suggest that these studies considerably underestimate the amount of target specificity that similar homeodomains can achieve in the context of the developing embryo.

4.1.4 Homeodomain Proteins as Transcriptional Activators.

As noted previously, homeodomains would appear to be sequence-specific DNA-binding regions of the proteins in which

they occur. Overall, these proteins are thought to play a role in differential gene expression. It is thought that they do this by acting as transcriptional activators, and evidence for this is discussed here.

A transcriptional activator, as the name implies, is a protein which serves to activate the transcription of a specific gene or genes. Such proteins seem to have a number of features in common. They possess both a DNA-binding surface and an activating region. The sequence-specific DNA-binding surface serves to position the protein on the DNA such that the other, activating surface, can interact with another protein (e.g. an RNA polymerase) to promote the initiation of gene transcription.

Some generalisations have been made about other properties of transcriptional activators, based upon studies of, for example, the yeast transcriptional activator, GAL4, and the prokaryotic repressors with the helix-turn-helix DNA-binding motif (Ptashne 1988; Maniatis et al.1987). Firstly, they tend to be modular proteins whose activating and DNA-binding regions are interchangeable. Also, whereas the DNA-binding surfaces are specific structures, each with a relatively high affinity for its cognate DNA-binding site, the activating regions are much less precisely defined structures characterized by an excess of acidic residues, although there has been a recent report of non-acidic activation regions in a eukaryotic transcriptional activator (Tora et al. 1989).

It is thought that a structural motif such as an α-helix is essential for the formation of an activating region (Ptashne 1988). It appears that an activator that works in yeast will also work in mammalian, plant and insect cells, provided that the target gene has an appropriate DNA-binding site in its promoter region (Ptashne 1988). Furthermore, any two of these activators can work cooperatively to stimulate transcription. There is variability

in the strength of the activating regions such that activators with stronger activating regions are capable of exerting their influence over greater distances on the DNA (Maniatis et al.1987). However, activators with stronger activating regions are likely to be tightly controlled (in terms of activating the activator) to limit the exposure of these regions in their fully functional form in the nucleus until their operation is required (Ptashne 1988).

Studies of mutants of the λ repressor protein, which has a dual repressor/activator function, have shown that the protein needs to bind to the DNA first in order to be positioned near RNA polymerase for the required protein-protein interactions (Ptashne 1988). Mutants which bind the DNA but fail to activate transcription have changes on the repressor surface in the region that most closely approaches the polymerase molecule. The activating surface of λ repressor seems to consist of an amphipathic α-helix bearing solvent-exposed residues which are predominantly negatively charged (Asp or Glu) (there are hydrophobic residues along the other surface) (Ptashne 1988). Mutants detrimental to the activating function reduce this negative charge. Both the P22 and 434 repressor also have this feature. In addition, λ Cro, which has the same DNA-binding specificity as λ repressor but does not activate transcription (Pabo and Sauer 1984), can be made into an activator by substituting acidic residues from λ repressor into the appropriate surface of Cro (Ptashne 1988).

Whilst λ repressor possesses both its DNA-binding and activating surfaces on the same domain, eukaryotic activators have them in different domains (Maniatis et al.1987; Ptashne 1988). Domain swaps in eukaryotic activators suggest DNA-binding is a "neutral" function, serving to bring the protein's activating surface to the vicinity of a gene, whilst the activating region can be in either orientation relative to the DNA-binding domain (Ptashne

1988).

Transcriptional activators from many organisms have been found to work cooperatively to significantly increase the levels of transcription. This has been interpreted as meaning that both activators simultaneously interact with a third protein, although this would suggest that there are specific protein-protein interactions, requiring specific variation in the activating region akin to that in the DNA-binding domain (Ptashne 1988).

The homeodomain-bearing proteins of *Drosophila* have been shown to display sequence-specific DNA-binding activity (as discussed earlier), yet their status as transcriptional activators *in vivo* is still not clear. However, three of the POU proteins were originally identified biochemically as mammalian transcriptional activators binding to identified DNA sequences before it was realised that they bore homeodomains. Pit-1 is a protein which is expressed exclusively in the pituitary gland where it activates the transcription of the prolactin and growth hormone genes. OCT-2 is expressed in lymphoid B cells and activates the transcription of immunoglobulin genes, whilst OCT-1 is a ubiquitous activator of the small nuclear RNA and histone H2B genes (He et al.1989). Both OCT-1 and OCT-2 have N-terminal blocks of charged amino acids and glutamines which could be activating regions (Sturm et al.1988).

It has been shown that both the *Ultrabithorax* and the *Abdominal-B* homeoproteins of *Drosophila* can activate transcription in mammalian cells, their activity and target specificity being similar to that of OCT-2 (Thali et al.1988). From this work, Thali et al. (1988) conclude that the proteins encoded by *Drosophila* homeotic genes are indeed transcriptional activators and that they can function in mammalian cells in the absence of any other *Drosophila* protein. All three homeoproteins are considered to be functionally equivalent, in that they stimulated

transcription by acting on the same set of target DNA sequences. In addition to a role as activators, genetic analyses have suggested that homeoproteins can act as repressors. Although it is possible that some of the observed interactions are indirect, there are precedents for eukaryotic regulators that can directly activate or repress transcription, depending on the target promoter. It is suggested that an accessory protein in *Drosophila*, which would also be a DNA-binding protein, could convert an activating homeoprotein into a repressing one by masking its activation domain (Thali et al.1988).

In another study, it was shown that *fushi tarazu*, *paired* and *zen* activated transcription by interacting with *engrailed* regulatory sequences, whilst the *even-skipped* and *engrailed* homeoproteins were found to repress, or quench, the activation induced by the other proteins (Han et al.1989). This work employed a transient expression assay in *Drosophila* tissue culture cells. A wide range of activities in the assay were found for the homeodomain proteins examined. An explanation for this is that it reflects variations in the protein binding to the five consensus binding sites used in the target sequence. However, Han et al. (1989) suggest that it is more likely that the range of assay activities are due to differences in the protein sequences outside the DNA-binding homeodomain. Attention is again drawn to a mutant lacking part of the C-terminal end of the *zen* protein that has lost its activating ability. The putative activating region contains 12 out of 64 acidic residues.

A further observation by Han et al.(1989) was that combinations of the homeoproteins tested resulted in substantial synergistic increases in expression. Since these different homeoproteins were recognising the same DNA sequence, a multi-switch model was proposed, wherein different combinations of homeoproteins assemble within the promoter of a target gene by

competing for multiple copies of a common binding site. The transcriptional activity of the target gene depended on exactly what combination of homeobox proteins was bound.

Further work on the *bcd* homeoprotein provides more evidence for activation of transcription *in vivo* in *Drosophila* (Driever and Nusslein-Volhard 1989). The *bcd* protein, as mentioned earlier, has a homeodomain which displays sequence-specific binding activity. The rest of the 489 amino acid protein has a number of features which are possibly significant to a role as a transcriptional activator. At the C-terminal end, there is an acidic region (16 acidic and 5 basic residues out of 67), and a glutamine-rich region which has been referred to as the M- or Opa-repeat (Wharton et al.1985). Both of these features have been noted a number of times in relation to transcriptional activators, and will be discussed later in the context of the *rough* protein. At the N-terminal end of *bcd*, there is also a histidine/proline-rich 30 amino acid sequence akin to that found in a number of other *Drosophila* genes. This is termed the Paired repeat and has uncertain function (McGinnis et al.1984b). It is also noted that *bcd* is rich in hydroxyl amino acids (37 serines, 27 threonines and 19 tyrosines), although the significance of this is unclear. There is evidence that *bcd* is subject to phosphorylation *in vivo* (Driever and Nusslein-Volhard 1989).

The *bcd* protein was shown to activate transcription in yeast and *Drosophila* tissue culture cells. However, Driever and Nusslein-Volhard (1989) also assayed the morphogenetic properties of *bcd* and various derivatives of it by injection of messenger RNAs into *Drosophila* embryos derived from *bcd*$^-$ mutant mothers. Thus, besides an assay for activation of a target gene, they could also demonstrate successful *bcd* function by rescue of the mutant embryos. Their results indicated that the N-terminal

half of the protein (246 amino acids), which includes the homeodomain, has a weak transcriptional activating function (as assayed in all systems), in comparison with the full 489 amino acid *bcd* protein, and it could only partially rescue the mutant embryos. However, when it was fused to certain acidic activating sequences from viral, yeast and *E. coli* transcriptional activators, the N-terminal *bcd* fragment fully rescued the mutant embryos.

They also noted that whilst the larger 396 and 348 amino acid N-terminal *bcd* derivatives could rescue and stimulate transcription in embryos, they had negligible activating activity in yeast. This may be due to modification of these proteins in the embryo, e.g. by phosphorylation, in order to potentiate or acquire an activating function. It is also possible that the glutamine-rich sequence from amino acids 247 to 348 functions as an activating sequence in the embryo but not in yeast or *Drosophila* tissue culture cells. Finally, they noted that there are other possible functions of *bcd* in the course of embryogenesis, such as inhibition of translation of maternal *caudal* mRNA or regulation of the domain of expression of another gene, *Kruppel*. If these are directly achieved by the *bcd* protein, then such functional activity must be specified in the N-terminal 246 amino acids.

However, the work by Kuziora and McGinnis (1989) provides the clearest evidence of transcriptional activation by homeoproteins in the *Drosophila* embryo. As described in section 4.1.3, they used a hybrid homeotic selector protein to demonstrate the target specificity of the homeodomain *in vivo*. Ectopic expression of the *Dfd* homeoprotein using a heat-shock inducible promoter activates the normal chromosomal copy of the *Dfd* gene, which is then persistently expressed in other segments via an autoregulatory circuit. The phenotypic result of the ectopic, persistent expression of *Dfd* is a transformation of head and thoracic segments towards a maxillary identity. Thus the

homeoprotein, *Dfd* , can activate transcription of its own expression in the *Drosophila* embryo. Yet, the chimeric *Dfd* protein into which the *Ubx* homeodomain had been substituted, failed to activate the chromosomal *Dfd* gene in this system, but was shown instead to activate ectopic transcription of the *Antp* gene in the embryo (Kuziora and McGinnis 1989).

The *rough* protein has a number of structural features similar to those associated with a function as a transcriptional regulator. The *rough* gene has 3 exons (Figure 1.1). The first, giving rise to the N-terminal domain, has a histidine/glutamine-rich region of 23 amino acids (10 His and 7 Gln) from residue 116 and an acidic region of 64 amino acids from residue 44 with 11 acidic, 4 basic and 20 hydrophobic residues. The second exon encodes the homeodomain, whilst the third, encoding the C-terminal domain of the protein, has as its most distinctive feature another glutamine-rich stretch of 17 amino acids (14 /17) (Tomlinson et al.1988; Figure 1.1).

The glutamine-rich regions have been noted in a number of other *Drosophila* proteins, and termed the M- (McGinnis et al.1984) or Opa repeat (Wharton et al.1985). Such Opa repeats have been found in other homeoproteins, including *bithorax*, *Antennapedia*, *engrailed*, and *bicoid*. A clear functional role for such glutamine-rich regions has yet to be ascribed although there is speculation as to a role in activation of transcription (Driever and Nusslein-Volhard 1989).

The most abundant amino acid in *rough* is glutamine (9.5%). However, the greater number of Opa-repeats make both glutamine (18%) and proline (10%) more abundant in the homeoprotein, *Antennapedia*, which is only a little larger (378 compared with 350 amino acids).

The proportion of hydroxyl amino acids in *rough* is similar to that found in *bcd* : 27 serines (7.7%), 22 threonines (6.3%) and 11

tyrosines (3.1%) versus 37 ser (7.6%), 27 Thr (5.5%) and 19 Tyr (3.9%). The significance of this is not examined by Driever and Nusslein-Volhard (1989). However, it is implied that there may be some processing of the *bcd* protein in the *Drosophila* embryo and that this might entail phosphorylation of the protein at hydroxyl groups.

In the absence of any direct evidence for the *rough* protein acting as a transcriptional activator, or that it is processed in order to facilitate such an activity, then one can only speculate as to the possible function of the regions of the *rough* protein on either side of the homeodomain.

4.1.5 The Promoter Region of the *rough* Gene.

There seem to be two DNA sequence elements which are required for the regulation of genes that encode mRNA in higher eukaryotes: promoters and enhancers (Maniatis et al.1987).

Promoters are located immediately upstream from the start of transcription and are usually 100 bp in length (Maniatis et al. 1987). The promoter is required for accurate and efficient initiation of transcription, whereas enhancers increase the rate of transcription from promoters. The distinctive characteristic of enhancers is that they can act on cis-linked promoters at great distances in an orientation-independent manner and can also function downstream from the transcription unit (Maniatis et al. 1987). A commom pattern of organisation of eukaryotic promoters has been identified: a typical promoter includes an AT-rich region designated the TATA box, and one or more sequence elements of 8 to 12 base pairs designated upstream promoter elements (UPE's, e.g. the CCAAT box). The TATA box functions primarily to ensure that transcripts are accurately initiated, whereas the UPEs increase the rate of transcription.

The relative strength of promoters is apparently

determined by the number and types of UPEs (Maniatis et al.1987). UPEs act regardless of their orientation with respect to the TATA box, although the distance between the UPEs and the TATA box, and their relative orientations on the DNA helix seem to affect the level of transcription, suggesting that one or more proteins bound to the UPE, interact with a protein (or proteins) bound to the TATA box, and that this interaction requires stereospecific alignment of the proteins on the DNA helix (Maniatis et al. 1987). UPEs and proteins that bind specifically to them have been delineated by *in vitro* DNA footprinting experiments and gel retardation assays.

Enhancers also contain discrete DNA sequence elements that specifically interact with proteins. They are around 40 bp long and can be located hundreds or sometimes thousands of nucleotides from the promoter, and in many cases, the major portion of the DNA is irrelevant to enhancer function (Ptashne 1986). There are a number of theories as to how regulatory proteins bound at an enhancer can influence the activity of another protein (notably RNA polymerase) located some distance away at the promoter. These include looping out of the intervening DNA ("looping"), changing the DNA conformation ("twisting"), moving of the protein along the DNA to another site ("sliding"), and the binding of a procession of proteins to the intervening DNA ("oozing") (Ptashne 1986). It has been shown that different proteins which bind to different sites within a single enhancer can stimulate transcription in specific cell types at different times (Maniatis et al.1987). Thus, these enhancer regions may be the sequences at which temporal and tissue-specific transcription factors bind to exert their influence.

The *rough* gene has a transcription unit of 4.3 kb which gives rise to a 1.3 kb mRNA encoding the 350 amino acid *rough* protein (Tomlinson et al.1988). The start of transcription is 64 bp upstream of the translation start codon, and has the RNA

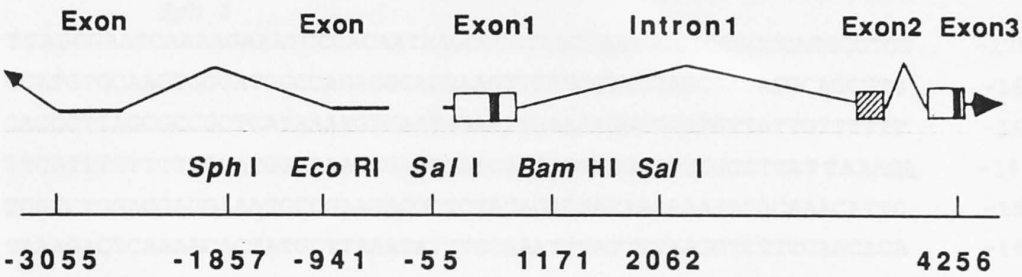


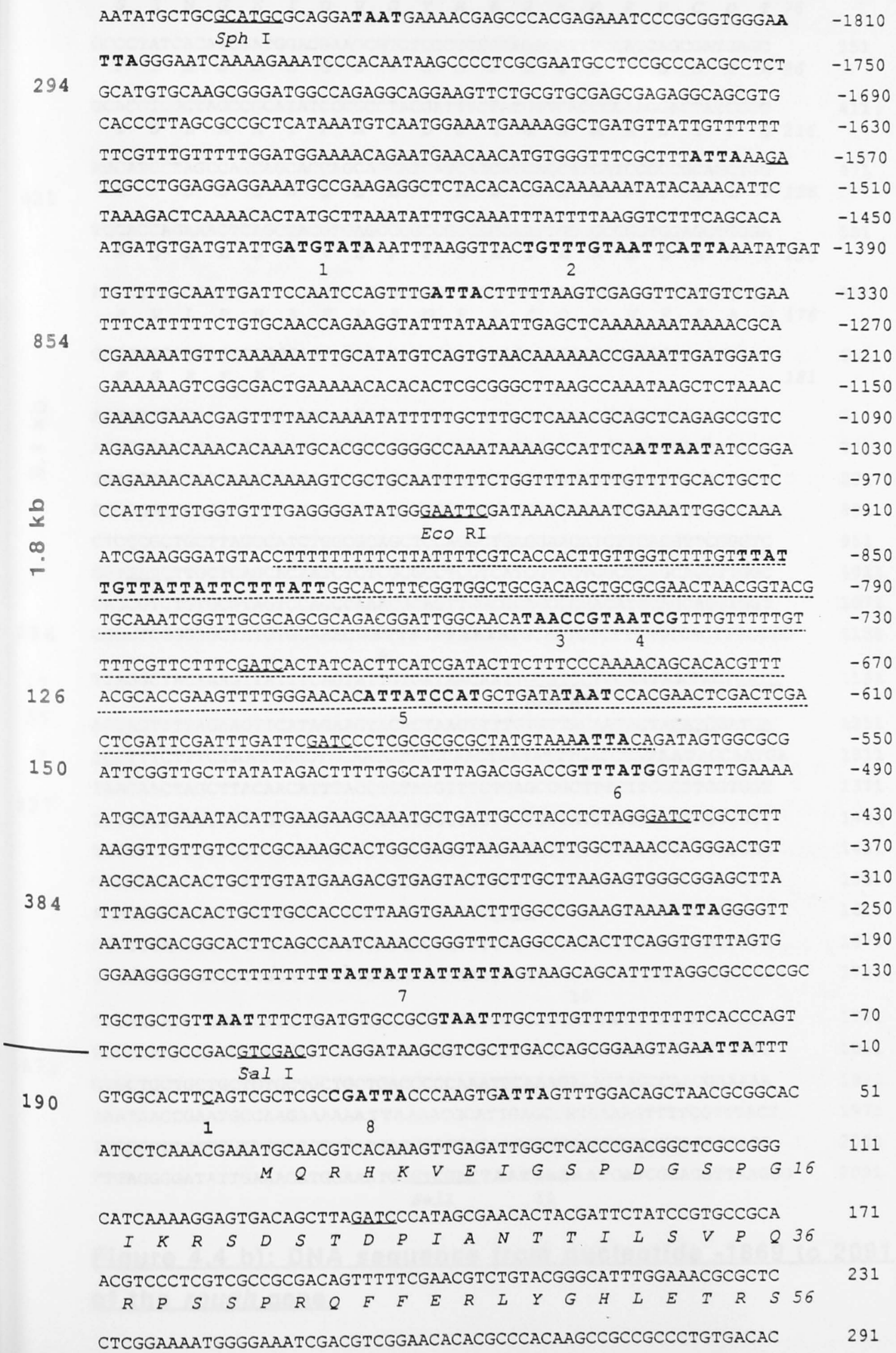
Figure 4.4 a): The structure of the *rough* locus, including 3kb upstream of the transcription start site (adapted from Tomlinson et al.1988).

The transcription unit of the *rough* gene is shown in more detail in Figure 1.1. There are two exons of another gene transcript, which is transcribed in the opposite direction (Tomlinson et al. 1988). The sequence from *Sph* I (-1857) to *Sal* I (2062) is given in Figure 4.4 b).

Figure 4.4 b): DNA sequence from nucleotide -1869 to 2091 of the *rough* gene (Tomlinson et al.1988).

The first exon of *rough* is designated by the one-letter amino acid code, whilst the first exon of the gene transcribed in the opposite direction (see above), is indicated by a dotted line (-562 to -894). Homeodomain consensus sequences (see Table 4.1) are shown in bold: 1,5, ATGNATA$_{T}^{A}$$_{T}^{A}$/$_{A}^{T}$$_{A}^{T}$TATNCAT; 2,3, TNNNNGTAAT; 4,8,9,10, TAATCG/CGATTA; 6, TTTATG; 7, $(TTA)_5$; 11, TAATGNNNNT. Many homeodomain consensus sequences contain the motif, TAAT/ATTA, and there are a number of copies of this motif throughout the sequence, which have been shown in bold. It is interesting to note that, whilst there are 22 TAAT/ATTA's in the 1900 bp 5' of *rough* exon1, and 27 within the first intron (607 to 3383), there are none in either the first or second exon of *rough*.

The restriction enzyme sites shown in Figure 4.4 a) are indicated. In addition, the *Sau*3A sites, GATC, have been underlined. The numbers in the right-hand margin indicate the size of the fragments resulting from *Sau*3A digestion of the 1.8 kb *Hind* III-*Sal* I (the *Hind* III site is present in the plasmid, pRO1, just 5' to the *Sph* I site at -1857), and the 2.1 kb *Sal* I-*Sal* I, pRO1 fragments (see Figure 4.11).


AATATGCTGCGCATGCGCAGGATAATGAAAACGAGCCCACGAGAAATCCCGCGGTGGGAA -1810
Sph I
TTAGGGAATCAAAAGAAATCCCACAATAAGCCCCTCGCGAATGCCTCCGCCCACGCCTCT -1750
294
GCATGTGCAAGCGGGATGGCCAGAGGCAGGAAGTTCTGCGTGCGAGCGAGAGGCAGCGTG -1690
CACCCTTAGCGCCGCTCATAAATGTCAATGGAAATGAAAAGGCTGATGTTATTGTTTTTT -1630
TTCGTTTGTTTTTGGATGGAAAACAGAATGAACAACATGTGGGTTTCGCTTTATTAAAGA -1570
TCGCCTGGAGGAGGAAATGCCGAAGAGGCTCTACACACGACAAAAAATATACAAACATTC -1510
TAAAGACTCAAAACACTATGCTTAAATATTTGCAAATTTATTTTAAGGTCTTTCAGCACA -1450
ATGATGTGATGTATTGATGTATAAATTTAAGGTTACTGTTTGTAATTCATTAAATATGAT -1390
1
2
TGTTTTGCAATTGATTCCAATCCAGTTTGATTACTTTTTAAGTCGAGGTTCATGTCTGAA -1330
TTTCATTTTTCTGTGCAACCAGAAGGTATTTATAAATTGAGCTCAAAAAAATAAAACGCA -1270
854
CGAAAAATGTTCAAAAAATTTGCATATGTCAGTGTAACAAAAAACCGAAATTGATGGATG -1210
GAAAAAAGTCGGCGACTGAAAAACACACACTCGCGGGCTTAAGCCAAATAAGCTCTAAAC -1150
GAAACGAAACGAGTTTTAACAAAATATTTTTGCTTTGCTCAAACGCAGCTCAGAGCCGTC -1090
AGAGAAACAAACACAAATGCACGCCGGGGCCAAATAAAAGCCATTCAATTAATATCCGGA -1030
CAGAAAACAACAAACAAAAGTCGCTGCAATTTTTCTGGTTTTATTTGTTTTGCACTGCTC -970
1.8 kb
CCATTTTGTGGTGTTTGAGGGGATATGGGAATTCGATAAACAAAATCGAAATTGGCCAAA -910
Eco RI
ATCGAAGGGATGTACCTTTTTTTTTTCTTATTTTCGTCACCACTTGTTGGTCTTTGTTTAT -850
TGTTATTATTCTTTATTGGCACTTTCGGTGGCTGCGACAGCTGCGCGAACTAACGGTACG -790
TGCAAATCGGTTGCGCAGCGCAGACGGATTGGCAACATAACCGTAATCGTTTGTTTTTGT -730
3
4
TTTCGTTCTTTCGATCACTATCACTTCATCGATACTTCTTTCCCAAAACAGCACACGTTT -670
126
ACGCACCGAAGTTTTGGGAACACATTATCCATGCTGATATAATCCACGAACTCGACTCGA -610
5
CTCGATTCGATTTGATTCGATCCCTCGCGCGCGCTATGTAAAATTACAGATAGTGGCGCG -550
150
ATTCGGTTGCTTATATAGACTTTTTGGCATTTAGACGGACCGTTTATGGTAGTTTGAAAA -490
6
ATGCATGAAATACATTGAAGAAGCAAATGCTGATTGCCTACCTCTAGGGATCTCGCTCTT -430
AAGGTTGTTGTCCTCGCAAAGCACTGGCGAGGTAAGAAACTTGGCTAAACCAGGGACTGT -370
ACGCACACACTGCTTGTATGAAGACGTGAGTACTGCTTGCTTAAGAGTGGGCGGAGCTTA -310
384
TTTAGGCACACTGCTTGCCACCCTTAAGTGAAACTTTGGCCGGAAGTAAAATTAGGGGTT -250
AATTGCACGGCACTTCAGCCAATCAAACCGGGTTTCAGGCCACACTTCAGGTGTTTAGTG -190
GGAAGGGGGTCCTTTTTTTTTATTATTATTATTATTAGTAAGCAGCATTTTAGGCGCCCCCGC -130
7
TGCTGCTGTTAATTTTCTGATGTGCCGCGTAATTTGCTTTGTTTTTTTTTTTCACCCAGT -70
TCCTCTGCCGACGTCGACGTCAGGATAAGCGTCGCTTGACCAGCGGAAGTAGAATTATTT -10
Sal I
190
GTGGCACTTCAGTCGCTCGCCGATTACCCAAGTGATTAGTTTGGACAGCTAACGCGGCAC 51
1
8
ATCCTCAAACGAAATGCAACGTCACAAAGTTGAGATTGGCTCACCCGACGGCTCGCCGGG 111
M Q R H K V E I G S P D G S P G 16
CATCAAAAGGAGTGACAGCTTAGATCCCATAGCGAACACTACGATTCTATCCGTGCCGCA 171
I K R S D S T D P I A N T T I L S V P Q 36
ACGTCCCTCGTCGCCGCGACAGTTTTTCGAACGTCTGTACGGGCATTTGGAAACGCGCTC 231
R P S S P R Q F F E R L Y G H L E T R S 56
CTCGGAAAATGGGGAAATCGACGTCGGAACACACGCCCACAAGCCGCCGCCCTGTGACAC 291

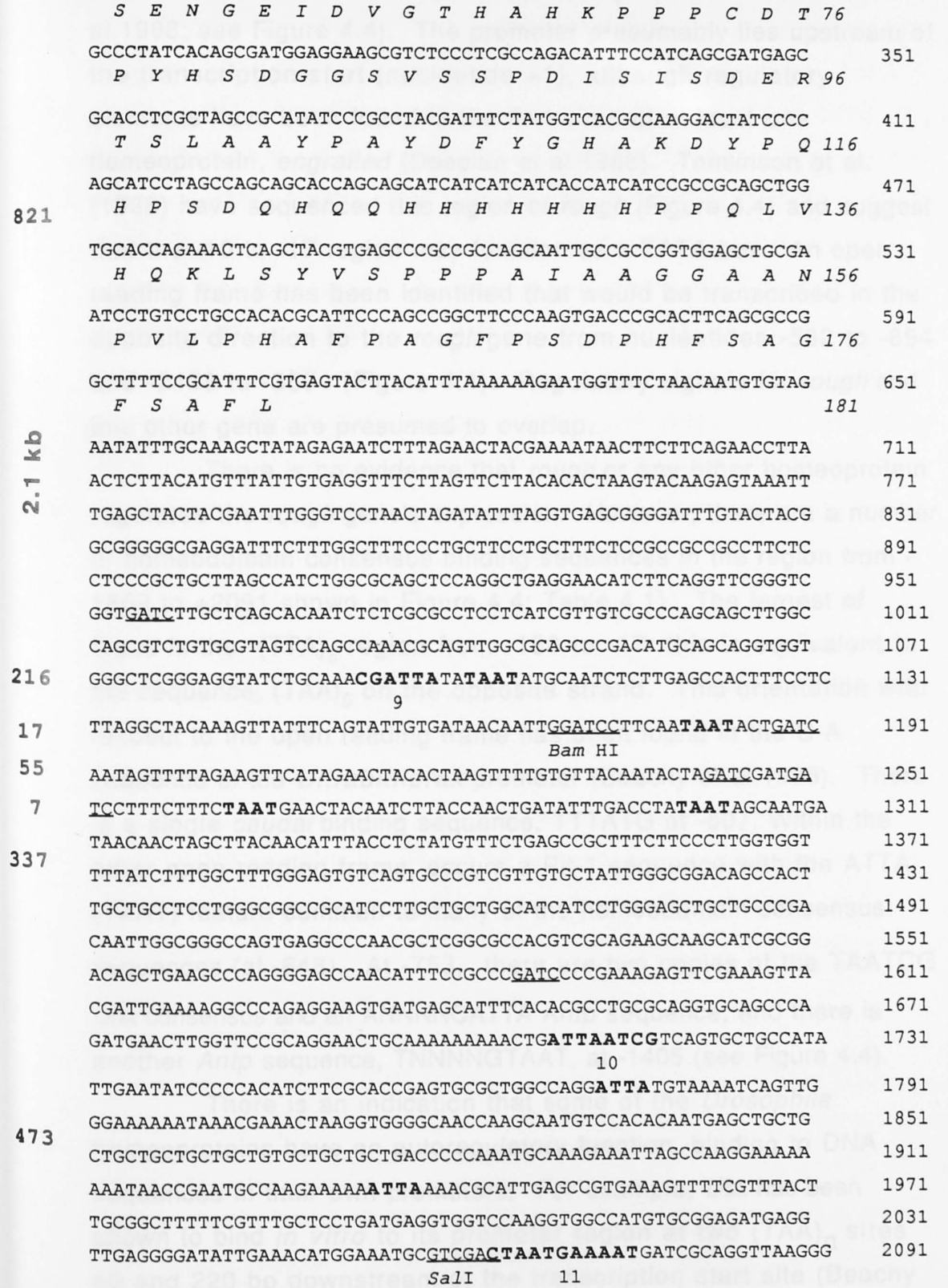


Figure 4.4 b): DNA sequence from nucleotide -1869 to 2091 of the *rough* gene.

polymerase II consensus sequence ($ATCA^{G}_{T}T^{C}_{T}$, Tomlinson et al.1988; see Figure 4.4). The promoter presumably lies upstream of the transcription start (nucleotide +1), although regulatory elements have been found in the first exon of at least one homeoprotein, *engrailed* (Desplan et al.1988). Tomlinson et al. (1988) have sequenced this region of *rough* (Figure 4.4) and suggest that the -10 to -17 region may function as a TATA box. An open reading frame has been identified that would be transcribed in the opposite direction to the *rough* gene from nucleotides -562 to -894 and -2458 to -3504 (Figure 4.4). Regulatory signals for *rough* and this other gene are presumed to overlap.

There is no evidence that *rough* or any other homeoprotein regulates the *rough* gene's expression. However, there are a number of homeodomain consensus binding sequences in the region from -1869 to +2091 shown in Figure 4.4; Table 4.1). The largest of these is the $(TTA)_5$ region from -156 to -17: this is equivalent to the sequence, $(TAA)_5$ on the opposite strand. This orientation with respect to the open reading frame has been found in the U-A sequence of the *Ultrabithorax* promoter (Beachy et al.1988). There is a single *caudal* binding sequence, TTTATG at -507. Within the other open reading frame, occurs a Pit-1 sequence with the ATTA (TAAT) feature common to many of the homeodomain consensus sequences (at -646). At -753, there are two copies of the TAATCG *Ubx* consensus and an ANNNNCATTA *Antp* sequence, and there is another *Antp* sequence, TNNNNGTAAT, at -1405 (see Figure 4.4).

There is an indication that some of the *Drosophila* homeoproteins have an autoregulatory function, binding to DNA sequences in their own promoters. For example, *Ubx* has been shown to bind *in vitro* to its promoter region at two $(TAA)_n$ sites 40 and 220 bp downstream of the transcription start site (Beachy et al.1988). It is thought that *Ubx* may have a positive feedback

loop in maintaining its expression *in vivo* during the later stages of embryogenesis. These studies also found that the TAA_4 motif was sufficient for *Ubx* binding. *Engrailed* binds to ten copies of the TCAATTAAAT consensus sequence in its own promoter *in vitro* : 7 of these sites were clustered between 600 and 900 bp upstream of the transcription start site, and the other three lie in the middle of the first intron (Desplan et al.1988). *Dfd* has been shown positively to autoregulate its own expression in some embryonic cells (Kuziora and McGinnis, 1989).

Thus, given that there is no indication which other gene or genes the *rough* protein may be regulating *in vivo*, it was decided to investigate the occurrence and nature of binding of the isolated *ro*HD to its own promoter region, in addition to the experiments with previously defined consensus sequences.

4.2 Results and Discussion.

4.2.0 Methods for Studying DNA-binding Interactions.

A number of methods have been developed for studying protein-DNA complexes. These have been used to provide biochemical and genetic data on the structural aspects of DNA recognition, its kinetics, thermodynamics, and the interaction of regulatory components (Hendrickson 1985).

The oldest of these methods is the nitrocellulose filter binding assay (Riggs et al 1970). Whereas double-stranded DNA will not bind to nitrocellulose during filtration, most proteins, and hence also any protein-DNA complexes, will be retained on the filters. Thus, by using radioactively labelled DNA fragments, and obtaining a measure of the radioactivity retained on the filters with a scintillation counter, one has an assay of the DNA-binding activity of a protein or protein extract (Hennighausen and Lubon

1987).

This method has been used to provide kinetic and equilibrium data for a number of sequence-specific protein DNA interactions, including the *lac* repressor-operator (Riggs et al 1970) and nuclear factor I (Hennighausen and Lubon 1987).

It has also been used as an assay for sequence-specific DNA binding proteins during their purification from cell protein extracts , for example, *Ultrabithorax* (Beachy et al.1988). There are some disadvantages associated with the use and accuracy of this technique, and these are discussed later in the context of results obtained using the *ro*HD.

A method which has gained increasing popularity in recent years as an alternative to the filter-binding assay is the mobility shift or gel electrophoresis-DNA binding assay (Fried and Crothers 1981; Garner and Revzin 1981; Hendrickson 1985; Sorger et al.1989). The essence of the method is simply to mix a DNA fragment containing a specific binding site with the protein of interest in an appropriate solution, including any necessary divalent ions and/or cofactors. After the binding reaction is complete, the sample is applied to a low ionic strength agarose or polyacrylamide gel and electrophoresed. The DNA distribution is detected by ethidium bromide staining and/or autoradiography which can be quantitated by densitometry. The resulting electrophoretic pattern will display a diminution or disappearance of the DNA band and the appearance of a lower mobility band containing the specific DNA-protein complex (Garner and Revzin 1986). This method has been used to provide quantitative measurements of protein-DNA interactions for many systems (Hendrickson 1985). These include the kinetics and equilibrium of the *lac* repressor-operator (Fried and Crothers 1981), RNA polymerase-*lac* promoter (Garner and Revzin 1981), and the heat shock transcription factor (Sorger et al.1989). It has also been

used to determine the stoichiometry of protein binding (Hendrickson 1985), the interactions of more than one protein during binding to a single DNA sequence (Hendrickson 1985), and as an assay for the purification of sequence-specific DNA-binding proteins from crude cell lysates (Hendrickson 1985, Sorger et al.1989; Scheidereit et al.1988). It has also been used to obtain data on the strength of binding (association/dissociation constants) for a number of homeodomain proteins, including *engrailed* and *fushi tarazu* (Desplan et al.1988), *even-skipped*, *zerknullt* and *paired* (Hoey and Levine 1988), the octamer transcription factors (OTF or Oct 1 and 2; Muller M. et al.1988; Scheidereit et al.1988), *Ultrabithorax* (Beachy et al.1988) and the *Antennapedia* homeodomain (Muller et al.1988), and was used as the purification assay technique for OTF-2 (Scheidereit et al.1988). The advantages and disadvantages of this technique, notably in comparison to the filter-binding assay are discussed later, in the context of results obtained with *ro*HD.

However, the most accurate definition of binding sites on the DNA is provided by the "footprinting" method (Galas and Schmitz 1978; Tullius 1989). Here, proteins bound to the DNA protect it from nuclease digestion (especially DNAase I protection; Galas and Schmitz 1978), or chemical modification. Fragments of the radiolabelled DNA are visualised by electrophoresis and autoradiography alongside the base-specific reaction products of the Maxam-Gilbert sequencing method (Galas and Schmitz 1978; Maxam and Gilbert 1977). It is then possible to see a protective "footprint" of the binding protein on the DNA sequence.

This method, unlike the previous two, permits the simultaneous monitoring of the extent of binding to multiple sites, since the electrophoresis gel on which the footprint pattern is displayed shows each site resolved from the others (Tullius 1989). Besides being used to identify DNA sequences to which a sequence-

specific protein will preferentially bind, the footprint titration experiment has been developed into a quantitative method for determining the energetics of protein binding to DNA (Tullius 1989). This typically entails a DNAase I protection reaction using a range of protein concentrations with fixed amounts of labelled DNA, the sites to which the protein binds most strongly will be protected at relatively lower protein concentrations. Equilibrium binding data for a number of homeodomain proteins have been obtained by footprinting, for example, *engrailed* and *fushi tarazu* (Desplan et al.1988), *eve*, *en*, *zen* and *prd* (Hoey and Levine 1988), and *Ubx* (Beachy et al.1988).

When working with an impure mixture of proteins, the technique of protein blotting can also be used to detect DNA-binding proteins (Bowen et al.1980; Hubscher 1987). This entails transferring proteins from a polyacrylamide gel to a nitrocellulose membrane as for Western blotting.

However, instead of probing the membrane with antiobodies, radiolabelled DNA fragments are used. Protein bands on the membrane capable of binding to the DNA will be shown by autoradiography. The technique has been used to detect *lac* repressor-operator interactions (Bowen et al.1980) and glucocorticoid receptor interactions with DNA (Silva et al.1987). It has the advantage of enabling one to identify the protein or proteins in the mixture which are capable of binding to the DNA.

4.2.1 Preliminary Experiments.

The initial experiments designed to investigate the DNA-binding ability of the *rough* homeodomain were conducted with the X17 fusion protein, whose construction is given in Chapter 5.

The DNA sequences which were used were from the *Ubx* promoter (3102 and $(TAA)_5$) and the engrailed promoter (NP_6).

3102 is a plasmid containing the entire *Ubx* promoter region as a 3 kb fragment, in which lies the sequence, 5'-TAAAAATAATAATAA- . TAA_5 and NP_6 are constructs obtained from C. Desplan (Desplan et al.1988) with consensus sequences present in the polylinker region of M13mp18. TAA_5 has the same consensus sequence as 3102, namely 5'-TAATAATAATAATAA, whilst NP_6 has 6 copies of the *engrailed* consensus sequence, 5'-TCAATTAAATGA. It was decided to test binding with these sequences since specific binding by homeoproteins of the *Antp* class had been shown for both, whilst a range of other homeodomains had also been shown to bind to the *engrailed* sequence (Table 4.1). The nearest prediction as to the likely specificity of the *ro*HD was on the basis of homology to the proposed recognition helix of the *Antp* class.

The first binding procedure tried was based on the South-Western blotting technique (Hubscher 1987). The X17 fusion protein extract was electrophoresed on SDS-PAGE, blotted from the gel onto a nitrocellulose membrane, and then immersed in solutions of end-labelled TAA_5, NP_6 and 3102 fragments. The membrane was rinsed and autoradiographed to show which protein bands had bound the labelled DNA. Thus, given that the protein extract was not pure, it was possible to see if it was capable of binding to any of the DNA fragments tested. Autoradiography showed that the fusion protein had bound fragments from all three solutions (not shown).

It was now decided to try to visualise this apparent binding by means of gel retardation to confirm that the fragments bearing the consensus sequences were indeed those which were being bound.

The initial experiments showed that there were some novel bands present in the X17-DNA lanes on the polyacrylamide gel. However, the TAA_5 and NP_6 M13 preparations gave background streaking on the gel, and the new band was rather diffuse. A much

clearer result was observed with 3102 - a new band was evident above the 273 bp band in the X17 lanes (not shown). However, although the 273 bp fragment bears the predicted binding sequence, the novel band may have come from other DNA fragments present in the digest of 3102.

It was decided to subclone the TAA_5 and NP_6 fragments into the polylinker region of the plasmid, pTZ18U. This pTZ DNA gave cleaner restriction enzyme digests with just the two fragments from the *Eco*RI/*Hind* III preparative digestion evident on gels.

A positive control against which to assess the extent of any DNA-binding was necessary. For this purpose, the plasmid pP_L*Ubx*Ib was obtained from Dr. L. Gavis (Stanford University) . This plasmid contains the *Ultrabithorax* cDNA transcribed from the P_L promoter. It has been used to produce *Ubx* protein for binding studies (Beachy et al.1988). With this plasmid, it was intended to obtain a *Ubx* protein extract with known affinity for both the TAA_5 and NP sequences. Thus, under the conditions at which the *Ubx* protein extract was capable of binding these fragments, one might expect to see equivalent activity from the *rough* derivatives.

Further binding studies of the *rough* homeodomain were conducted with the *ro*HD preparation from pJG(2), described in Section 3.2.4.

4.2.2 Gel Retardation Studies with the *rough* homeodomain.

For these experiments the 67 bp TAA_5 and 125 bp NP_6 *Eco*RI/*Hind* III fragments bearing the consensus sequences were purified on acrylamide gels. Changes in the migration of a single band could now be obtained.

The *ro*HD preparation had an estimated concentration of 0.42 µg/µl or 6.3×10^{-5} M.

Figure 4.5: Autoradiogram showing the electrophoretic mobility of NP_6 and TAA_5 DNA fragments through a native polyacrylamide gel with increasing concentrations of the *rough* homeodomain.

Binding reactions: [α-^{32}P]dATP end-labelled DNA fragments (67 bp TAA_5, and 125 bp NP_6) were mixed with varying amounts of *ro*HD in binding buffer (20 mM Tris pH7.5, 75 mM NaCl, 1 mM DTT, 10% glycerol) to give a total reaction volume of 40 µl. Reactions were incubated on ice for one hour and then electrophoresed.
Reactions: NP_6(58 nM) / TAA_5(125 nM). [*ro*HD] = **A**, 0 nM; **B**, 42 nM; **C**, 420 nM; **D**, 1050 nM; **E**, 2100 nM.

Gel electrophoresis: 15% 0.5 x TBE polyacrylamide gel was pre-electrophoresed at 300 volts (~18 mA) for 2 hours at 4°C, before loading the binding reactions. Electrophoresis after loading was at 300 volts (~18 mA) for 3 hours at 4°C. The gel was then dried and autoradiographed.

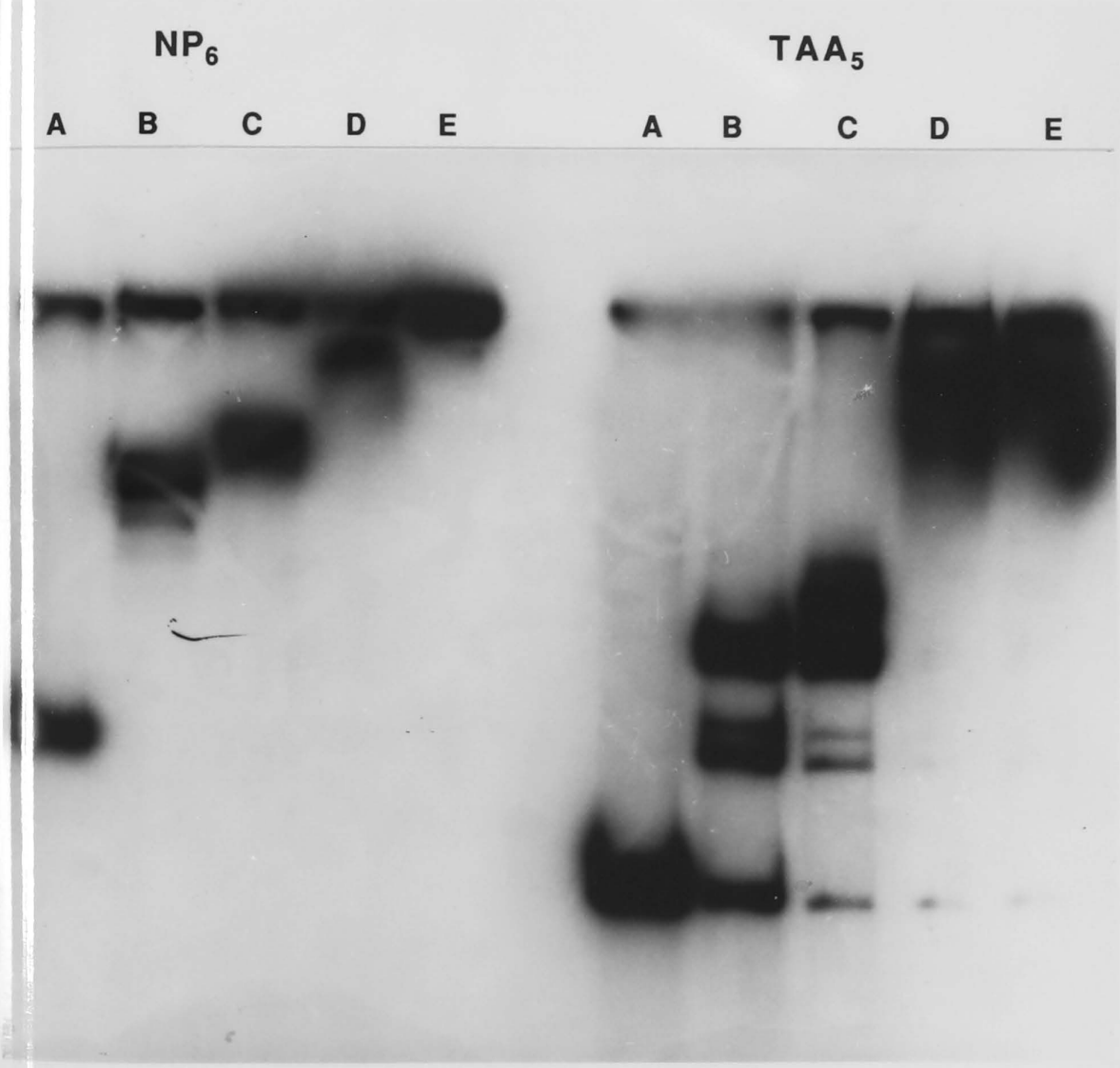


Figure 4.5: Autoradiogram showing the electrophoretic mobility of NP_6 and TAA_5 DNA fragments through a native polyacrylamide gel with increasing concentrations of the *rough* homeodomain.

Although it had been intended to use pure *Ubx* protein, it emerged that the apparent yields of the protein as visualised on SDS-PAGE were perhaps even lower than those obtained for the *ro*HD. However, since crude extracts had been successfully used in filter binding assays (Beachy et al.1988), these were tested for activity.

The lifetime of protein-DNA complexes can be profoundly affected by electrophoresis conditions, e.g. DNA gyrase will only form stable DNA complexes in the presence of 5 mM Mg^{2+} (Garner and Revzin 1986). Thus since no one set of conditions is appropriate for all systems, it is necessary to optimise them for each system. In the preliminary retardation experiments, a range of binding reaction buffers, and gel buffers, were tried, all of them based on Tris at pH 7 to 8, with various other ingredients. With the publication of the paper showing specific binding by the *Antp*HD (Muller et al.1988), it was decided to emulate the conditions contained therein, with 15% polyacrylamide gels, 0.5 x TBE electrophoresis buffer, and a similar reaction buffer (20mM Tris 7.5, 1mM DTT, and 10% glycerol, with an initial [NaCl] of 75mM, although this was increased to test the stringency of binding). Binding reactions had been conducted at temperatures upto 30°C however, the *Antp* work incubated the reaction on ice for 30 minutes before loading onto the pre-electrophoresed gel. Electrophoresis was at 300 volts at 4°C.

The initial study of the *ro*HD's binding capacity was with fixed amounts of TAA_5/NP_6 and variable concentrations of *ro*HD (0 to 500 nM) at 75mM NaCl. The reaction proceeded for 30 minutes, before the binding mixtures were loaded on the gel. This experiment showed an increasing level of retardation of both fragments with increasing [*ro*HD] (see Figure 4.5). The retardation of the NP_6 fragment seemed to be "cleaner" with fewer bands and a

greater relative retardation at each *ro*HD concentration, whilst the TAA_5 fragment formed a range of protein-DNA complexes for each treatment. This may be due to non-specific binding of the *ro*HD to the DNA fragments Given the larger size of the NP_6 band, one would expect a higher degree of such interaction with it rather than with the TAA_5 band. Alternatively, it may be due to specific binding - with 6 consensus sequences, NP_6 might be more tightly bound than the single-sequence TAA_5 fragments.

As discussed in Section 4.1.1, the strength of protein interactions with the DNA depends not only upon the concentration of protein and DNA, but also on the concentration of charge-neutralising cations in the solution. Thus, given the satisfactory degree of retardation obtained with 420nM *ro*HD, the stringency of binding was tested by raising the [NaCl] to 500mM. The retardation of both NP_6 and TAA_5 was maintained, with NP_6 showing a fixed band from low to high salt, whilst TAA_5 showed a change in the distribution of the complexes formed, smaller ones predominating at the higher salt concentration (not shown). This implies that some of the complexes formed with TAA_5 were less tightly bound, but that overall the *ro*HD forms relatively strong complexes with both of these fragments.

It was decided to further investigate the specificity of binding under these conditions by comparing the *ro*HD with some other proteins. These were the partially purified *Ubx* preparation (see Methods), the bacterial DNA-binding protein, HU (purified in Section 3.2.2), and egg-white lysozyme (Sigma). Varying concentrations of these protein solutions and *ro*HD were mixed with TAA_5 and NP_6, and electrophoresed as above. All of the proteins retarded the motion of the fragments through the gel (not shown). Indeed, at higher concentration of protein, the DNA fragments barely left the loading well, although this phenomenon has been ascribed to the aggregation of protein which can also

Figure 4.6: Autoradiogram showing changes in the elctrophoretic mobility of the NP_6 and TAA_5 DNA fragments by the *rough* homeodomain through a native polyacrylamide gel with increasing concentrations of NaCl.

Binding reactions: [α-^{32}P]dATP end-labelled DNA fragments (43 nM TAA_5 / 20 nM NP_6) were mixed with *ro*HD (420 nM) in binding buffer (20 mM Tris pH7.5, 1 mM DTT, 10% glycerol) containing non-specific competitor DNA (1 µg pCE30 = 15 times [TAA_5/NP_6]), and increasing concentrations of NaCl, to give a total reaction volume of 40 µl. Reactions were incubated on ice for 30 minutes and then electrophoresed.

[NaCl] = **A**, 60 mM (0 nM *ro*HD); **B**, 60 mM; **C**, 120 mM; **D**, 240 mM; **E**, 360 mM; **F**, 480 mM; **G**, 600 mM; **H**, 720 mM.

Gel electrophoresis: 10% 0.5 x TBE polyacrylamide gel was pre-electrophoresed at 300 volts for 1 hour at 4°C. After loading the binding reactions, electrophoresis was at 300 volts (~18 mA) for 2 hours at 4°C. The gel was then dried and autoradiographed.

Figure 4.6: Autoradiogram showing changes in the electrophoretic retardation of the TAA_5 and NP_6 DNA fragments by the *rough* homeodomain through a native polyacrylamide gel with increasing concentrations of NaCl.

result in the smearing of complexes through the gel (Hendrickson, 1985). Thus, under these conditions, there is evidently scope for a fair degree of binding which is presumably of a non-specific nature.

To further eliminate non-specific interactions, unlabelled competitor DNA was added in addition to the use of higher salt concentrations. Uncut pCE30 was chosen as the competitor DNA since it contained neither of the consensus sequences (the nearest to TAA_5 is 11 out of 15 bp and to the NP sequence is 8 bp out of 10). With the addition of 1μg of of pCE30 at 500mM NaCl, only the *rough* homeodomain and *Ubx* proteins bound to both TAA_5 and NP_6 (not shown). There was still some binding evident at 75mM NaCl by HU and lysozyme, however this disappeared completely at 500mM NaCl.

The more specific binding of the *ro*HD to TAA_5 and NP_6 in the presence of excess non-specific DNA was investigated at varying [NaCl] up to 720 mM (see Figure 4.6). This again shows what appears to be tighter binding to NP_6, since the main complex formed is stable at > 600 mM NaCl, whilst there are multiple bands evident in the TAA_5 lanes throughout, binding being completely eliminated at 720 mM NaCl (see Figure 4.6). However, there is some indication that the strength of binding to NP_6 is associated with the multiple copies of the NP sequence since there are six bands (at >600 mM NaCl) perhaps representing *ro*HD-NP_6 complexes of sequentially varying stoichiometry.

4.2.3 Binding Studies with a 15-mer TAA_5 oligonucleotide.

There was a possibility that the specific binding of the *ro*HD to the TAA_5 and NP_6 fragments observed in the previous section may have been influenced by the sequence from the M13 polylinker on either side of the consensus regions. This was investigated with an oligonucleotide containing just the TAA_5 consensus sequence.

Figure 4.7: Autoradiogram showing changes in the electrophoretic migration of 15-mer TAA_5 and mismatched TAA_5 oligonucleotides in the presence of the *rough* homeodomain and *Ubx* Ib extracts.

Binding reactions: The [γ-^{32}P]dATP labelled TAA_5 15-mer oligonucleotide, and mismatched, end-filled, oligonucleotides (see Section 4.2.3), were mixed with either *ro*HD (420 nM) or *Ubx* Ib extract (3 µl) in binding buffer (20 mM Tris pH7.5, 1 mM DTT, 10% glycerol) containing non-specific competitor DNA (1 µg pCE30), and increasing concentrations of NaCl, to give a total reaction volume of 40 µl. Reactions were incubated on ice for 30 minutes and then electrophoresed.

Reactions: *ro*HD (420 nM) + 15-mer TAA_5 oligonucleotide (130 nM): [NaCl] = **A**, 60 mM (0 nM *ro*HD); **B**, 60 mM; **C**, 240 mM; **D**, 480 mM; **E**, 720 mM. *Ubx* Ib extract (3 µl) + 15-mer TAA_5 oligonucleotide (130 nM): **F**, 60 mM; **G**, 240 mM; **H**, 480 mM; **I**, 720 mM.
*ro*HD (420 nM) + mismatched TAA_5 oligonucleotides (2 µM): **J**, 60 mM (0 nM *ro*HD); **K**, 60 mM; **L**, 480 mM. *Ubx* Ib extract (3 µl) + mismatched TAA_5 oligonucleotides (2 µM): **M**, 60 mM; **N**, 480 mM.

Gel electrophoresis: 15% 0.5 x TBE polyacrylamide gel was pre-electrophoresed at 300 volts for 1 hour at 4°C. After loading the binding reactions, electrophoresis was at 300 volts (~18 mA) for 2 hours at 4°C. The gel was then dried and autoradiographed.

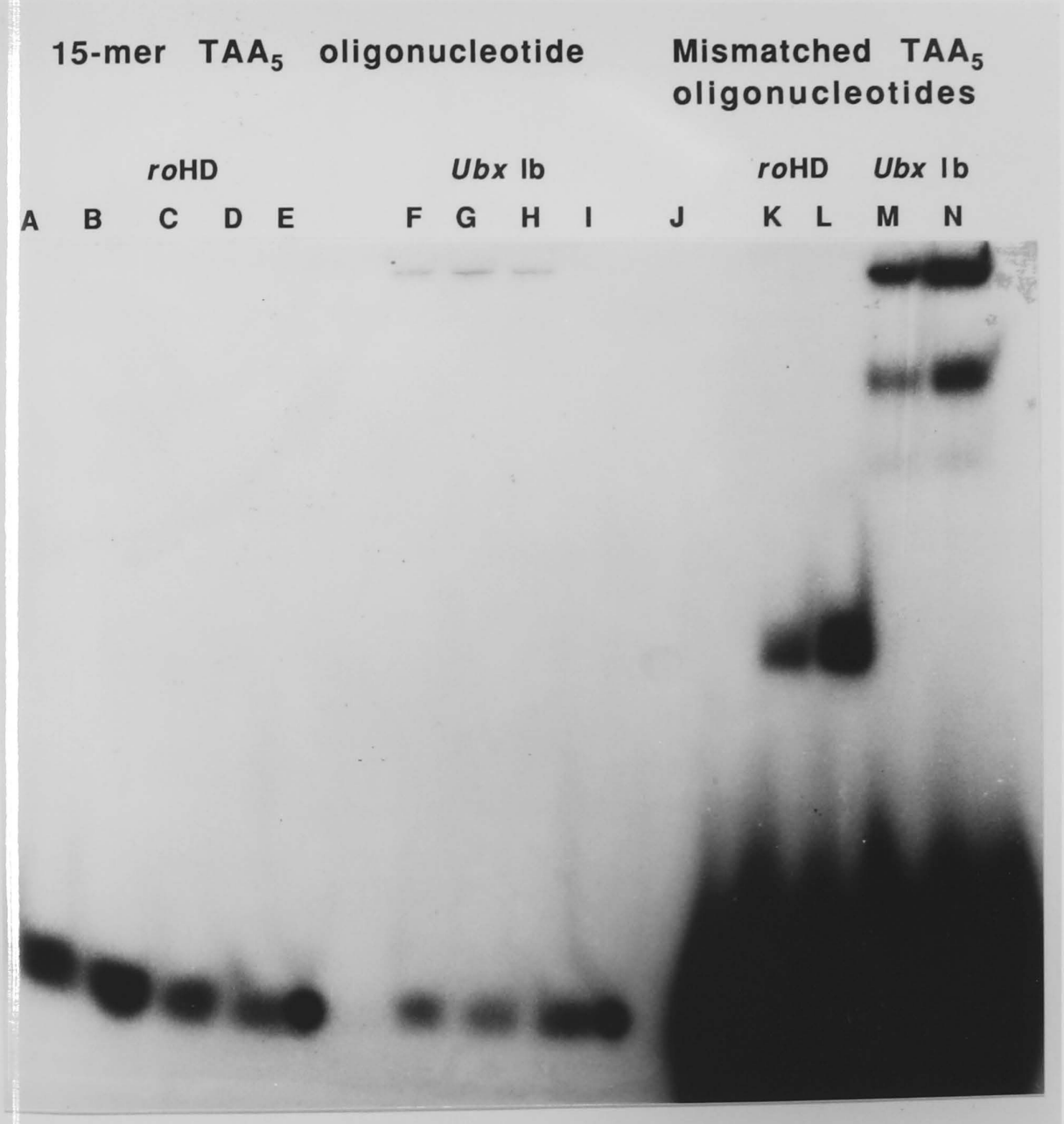


Figure 4.7: Autoradiogram showing changes in the electrophoretic migration of 15-mer TAA_5 and mismatched TAA_5 oligonucleotides in the presence of the *rough* homeodomain and *Ubx* Ib extracts.

It had also been intended to test the NP sequence (12 bp) and the *bicoid* sequence (9 bp, TCTAATCCC), the latter being a sequence for which the *ro*HD should have no specificity (Table 4.1; Figure 4.2). However, it was not known what minimum size of oligonucleotide was necessaary for binding. The *Ubx*Ib protein had been found in a "footprinting" experiment to be capable of binding to a minimum of a TAA_4 sequence in a 21 bp fragment (Beachy et al.1988), and the *Antp*HD bound a 26 bp oligonucleotide (Muller et al.1988). Thus, the largest of the consensus sequences, TAA_5, was synthesized as a 15-mer.

This oligonucleotide was labelled by phosphorylating the 5' end with polynucleotide kinase and [γ-^{32}P] ATP. The binding of the *ro*HD to the TAA_5 oligonucleotide was investigated using gel retardation and by filter binding. A range of [*ro*HD] were tried, with and without non-specific competitor DNA (1 µg pCE30), at upto 720 mM NaCl. There was no evidence of any binding by the *ro*HD in any of these studies. No novel band was evident under the conditions which resulted in binding to the 67 bp TAA_5 fragment (see Figure 4.7); whilst filter binding with oligonucleotide (4000 c.p.m. per reaction) showed that there was no ^{32}P retained.

However, when the *Ubx*Ib extract was tested, there was some oligonucleotide retarded at [NaCl] up to 500 mM, although the amount retarded was small relative to that observed with the 67 bp TAA_5 (Figure 4.7). Given the impure nature of the *Ubx* extract, this apparent binding might be due to the activity of proteins other than *Ubx* .

The inability of the *ro*HD to bind to the oligonucleotide may have been because the 15-mer was too short to allow adequate interaction between the DNA and the protein. This possibility was investigated by generating larger TAA sequences. These were made by heating a mixture of the single-stranded oligomers, 5'-TAA_5 and

5'-TTA_5 to 65°C, and then rapidly cooling them on ice to generate mismatched fragments with TAA and ATT overhangs at each end, e.g. 5'-TAA$T_A A_T A_T T_A A_T T_A A_T T_A A_T T_A A_T$TATT-5'. These fragments were labelled by end-filling to give 18-mers (as with the example), 21-mers, 24-mers, and perhaps even 27-mers. This mixture was used in gel retardation studies with both the *ro*HD and *Ubx* extracts at 75 and 500 mM NaCl. Retardation was now clearly evident (Figure 4.7) with both the *ro*HD and *Ubx*, even at 500 mM NaCl in the presence of non-specific competitor DNA (20 times excess). Thus, it appears that whilst the *ro*HD does bind specifically to the $(TAA)_n$ sequence (as was suggested with the 67 bp fragment), a minimum fragment size >15 bp is necessary to achieve a binding interaction which is sufficient to be detected by the techniques used. DNase I "footprinting", being a more sensitive method of detecting sequence-specific DNA-binding interactions, might show the affinity of the *ro*HD for these sequences more clearly.

It was not possible to identify which of the larger TAA oligonucleotides was bound, although this could be investigated by purifying the various end-labelled mismatched oligonucleotides and then checking them for binding. The binding of the *ro*HD to the TAA_5 sequence is presumably limited by the fragment size rather than the sequence length. Competitor DNA fragments with M13 polylinker sequence, and around the same size as the TAA_5/NP_6 pTZ fragments were tested in a quantitative assay which showed a clear preference for the consensus fragments (see next section.)

4.2.4 Quantitative Measurements of *ro*HD-DNA binding

Biochemical constants for protein-binding interactions can be obtained by a number of methods using the techniques of filter-binding, gel retardation or "footprinting", as discussed

earlier (section 4.2.1). This section describes the use of both the nitrocellulose filter-binding assay, and gel retardation method to obtain data on the strength of the binding of the *ro*HD to the TAA_5 and NP_6 fragments.

A standard measure of the strength of protein-DNA interaction is the equilibrium constant. For a simple interaction of protein and DNA in solution, the reaction may be presented as:-

$$[P] + [D] \underset{K_d}{\overset{K_a}{\rightleftharpoons}} [PD]$$

where [P] is the initial protein concentration, [D] is the DNA concentration, and [PD] is the concentration of the protein-DNA complexes formed. This assumes a simple stoichiometry for the reaction and is derived from the Law of Mass Action (e.g. Riggs et al.1970). This reaction reaches an equilibrium position in which the extent to which PD is formed from P and D is given by the association constant, K_a, and the degree to which PD dissociates to give its components is given by the dissociation constant, K_d.

At equilibrium, $K_a = \frac{[PD]_e}{[P]_e[D]_e}$ $\qquad$ $K_d = \frac{[P]_e[D]_e}{[PD]_e}$

the subscript e, indicating that these are the concentrations of the species at equilibrium. Thus, $K_d = 1/K_a$. The strength of protein-DNA binding is usually given in terms of the dissociation constant, K_d.

The equation (*): $[P]_{0.5} = K_d + 0.5\,[D]$ may be derived from the above equations (Riggs et al.1970) where $[P]_{0.5}$ is the concentration of protein which gives a half-saturation of the DNA.

The dissociation constant may be obtained by producing a standard binding curve with the DNA concentration held constant, usually at a level well below the concentration of the protein, which is varied within a moderate range spanning the estimated K_d

(Hendrickson 1985). Under these conditions the free protein concentration is essentially equal to that of the input protein, and the equilibrium binding constant is equal to the concentration of active protein at which 50% of the DNA is bound (at higher [D], the equation * may be used).

The fraction of the added protein that is active (that is, capable of binding to the DNA), the rate of approach to equilibrium, and other factors such as the possible association of protein subunits should be considered when calculating the K_d (Hendrickson 1985). A number of methods may be used to determine the amount of active protein in a protein preparation (Riggs et al.1970), for example, by titrating the protein sample at a concentration of DNA fragment well above the expected dissociation constant (~100 times). The rate of approach to equilibrium, and thus the length of time necessary to incubate the binding reaction in order to attain equilibrium, may be obtained by studying the time course of binding.

With the *ro*HD preparation used, it was assumed that the protein was fully active, whilst a simple time course of binding at 4°C and 30°C suggested that equilibrium is reached within 5 minutes at both temperatures (not shown).

Standard binding curves for reactions betwen the *ro*HD at varied concentrations, and fixed concentrations of DNA fragments, were generated using both gel retardation and the filter-binding assay (Figure 4.8).

It emerged that these techniques gauge consistently different results under the same conditions. The binding detected by gel retardation occurred at lower *ro*HD concentrations than was shown by the filter binding assay, and hence estimates of the dissociation constant derived by gel retardation are an order of magnitude lower (i.e. indicating stronger binding) than those

Figure 4.8 a),b),c),d): Graphs showing binding of DNA fragments with [*ro*HD] a)TAA_5, b)NP_6, c)82 bp, d)132 bp.

Binding reactions: [α-^{32}P]dATP end-labelled DNA fragments were mixed with varying amounts of *ro*HD in binding buffer (20 mM Tris pH7.5, 75 mM NaCl, 1 mM DTT, 10% glycerol) to give a total reaction volume of 30 µl (gel retardation) or 40 µl (filter-binding). Reactions were incubated at room temperature (25°C) for 30 minutes, and then electrophoresed or passed through a nitrocellulose filter.

Reactions: a) **TAA_5** = 1 nM (gel)/5 nM (filter), [*ro*HD] = 0.55 to 1120 nM (gel)/4.2 to 420 nM (filter); b) **NP_6** = 0.44 nM (gel)/2.2 nM (filter), [*ro*HD] = 0.55 to 1120 nM (gel)/4.2 to 420 nM (filter); c) **82 bp** = 1 nM (gel)/5 nM (filter), [*ro*HD] = 0.76 to 1495 nM (gel)/4.2 to 2100 nM (filter); d) **132 bp** = 1 nM (gel)/5 nM (filter), [*ro*HD] = 0.76 to 1495 nM (gel)/4.2 to 2100 nM (filter).

Gel electrophoresis: Method as in Figure 4.10. Densitometry as in Section 4.3.7.

Filter-binding assay: Method as in Section 4.3.6.

Table 4.2: Equilibrium binding results from Figure 4.8.

	Gel retardation	Filter binding	Filter : gel result
TAA_5	8.8×10^{-9}M	2.4×10^{-7}M	27
NP_6	9.8×10^{-9}M	1.7×10^{-7}M	17
82 bp	1.9×10^{-7}M	$>2.1 \times 10^{-6}$M	>11
132 bp	5.9×10^{-7}M	$>2.1 \times 10^{-6}$M	> 4

Calculated using the equation: $[P]_{0.5} = K_d + 0.5\,[D]$

% unbound TAA_5

% unbound NP_6

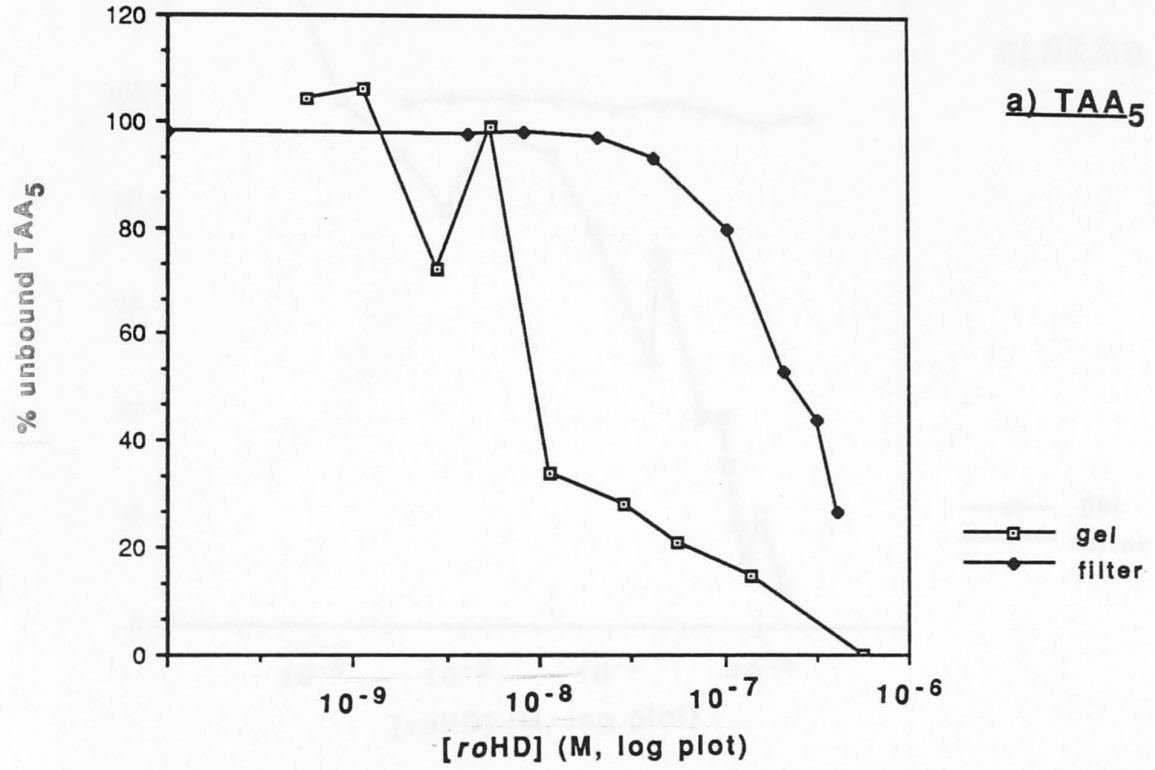
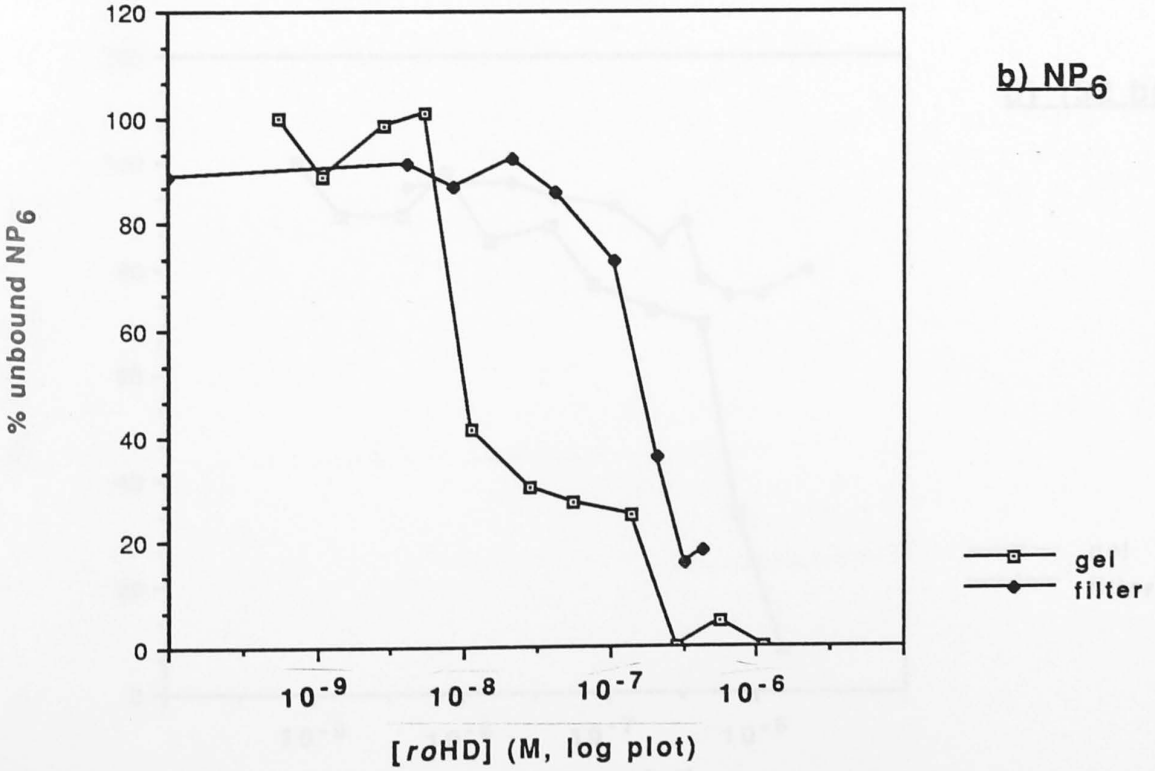


Figure 4.8: Graphs showing binding of DNA fragments with [*ro*HD]: a) TAA_5; b) NP_6.



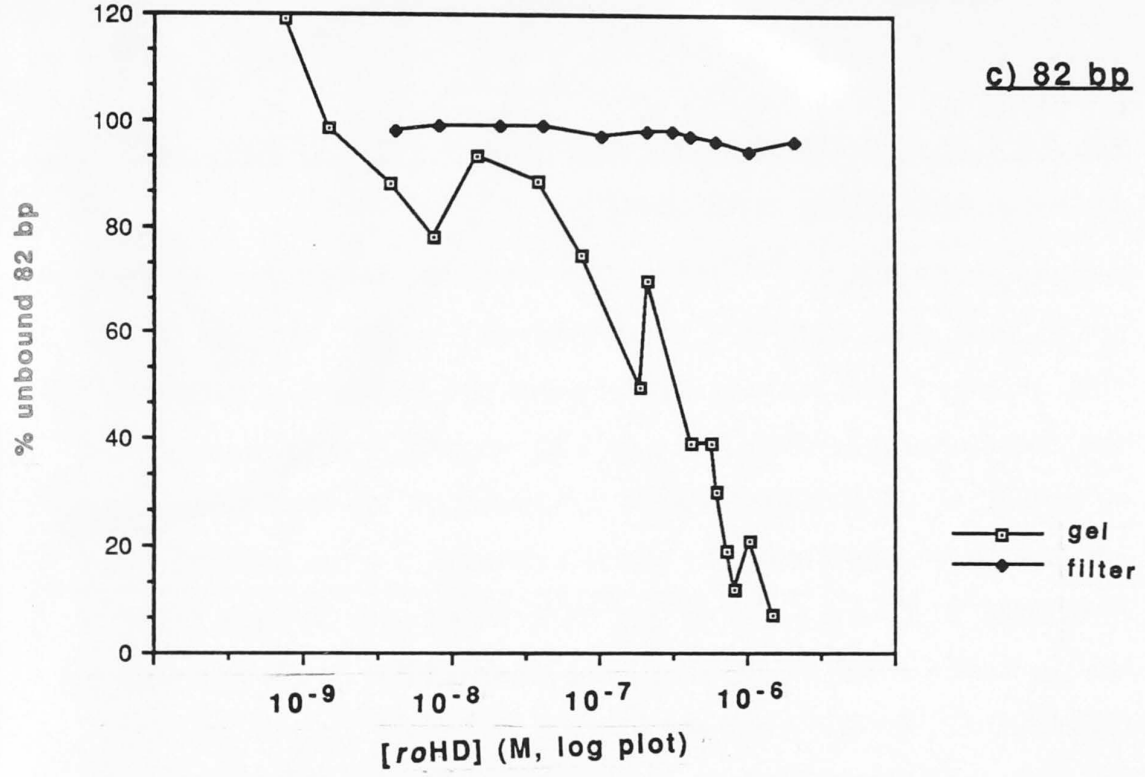


Figure 4.8: Graphs showing binding of DNA fragments with [*ro*HD]: c) 82 bp; d) 132 bp .

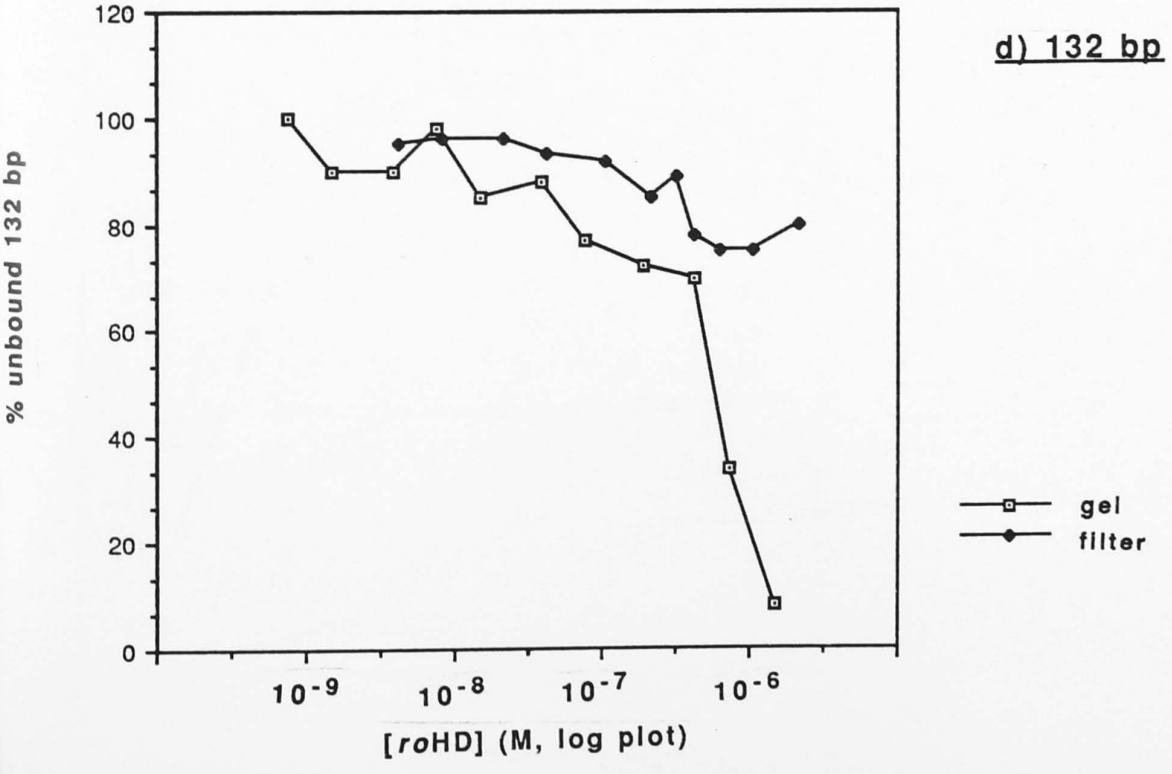


Figure 4.9: Graphs showing changes in *ro*HD binding of a)TAA_5, and b)NP_6 with [NaCl].

Binding reactions: [α-^{32}P]dATP end-labelled DNA fragments were mixed with *ro*HD in binding buffer (20 mM Tris pH7.5, 1 mM DTT, 10% glycerol) with varying [NaCl], with (**C**), or without (**NC**), non-specific competitor DNA (1 µg pCE30) to give a total reaction volume of 40 µl. Reactions were incubated on ice for 30 minutes, and then electrophoresed or passed through a nitrocellulose filter. Reactions: a) **TAA_5** = 71 nM (gel NC), 43 nM (gel C)/6 nM (filter NC/C), [*ro*HD] = 420 nM; b) **NP_6** = 34 nM (gel NC), 20 nM (gel C)/3 nM (filter NC/C), [*ro*HD] = 420 nM.

Gel electrophoresis: Method as in Figure 4.10. Densitometry as in Section 4.3.7.

Filter-binding assay: Method as in Section 4.3.6.

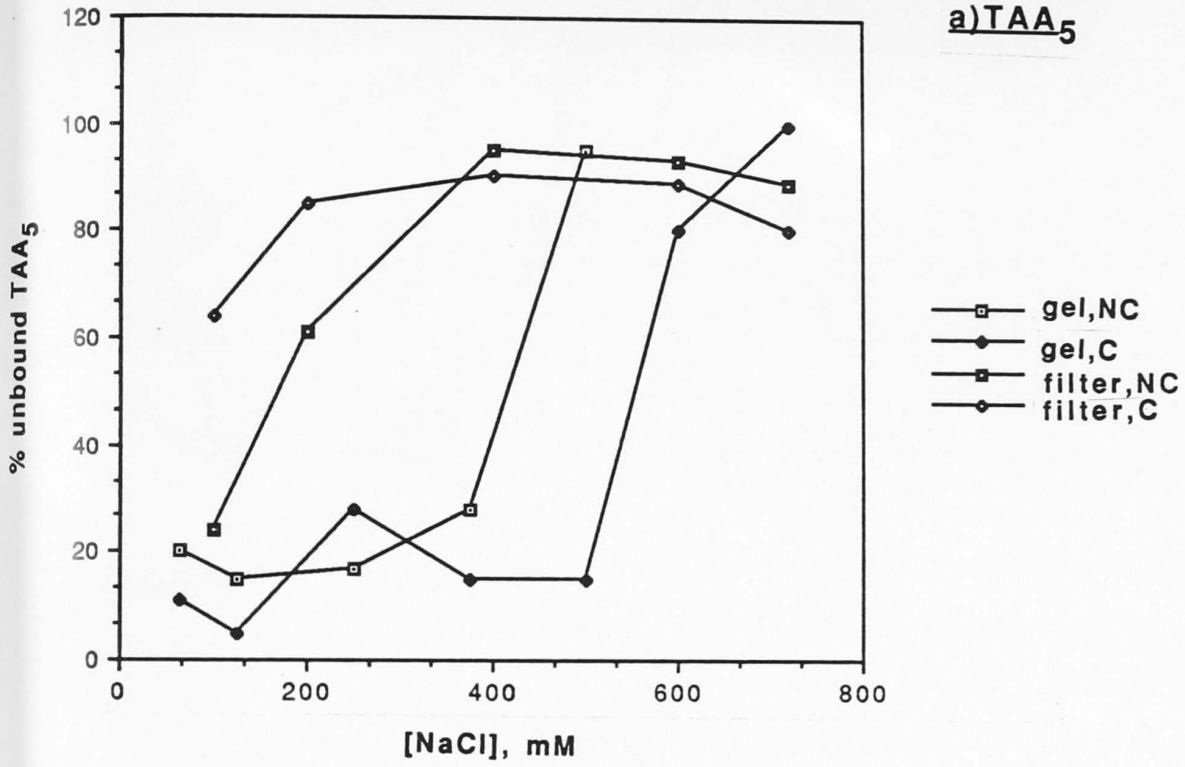


Figure 4.9: Graphs showing changes in *ro*HD binding of a)TAA_5, and b)NP_6 with [NaCl].

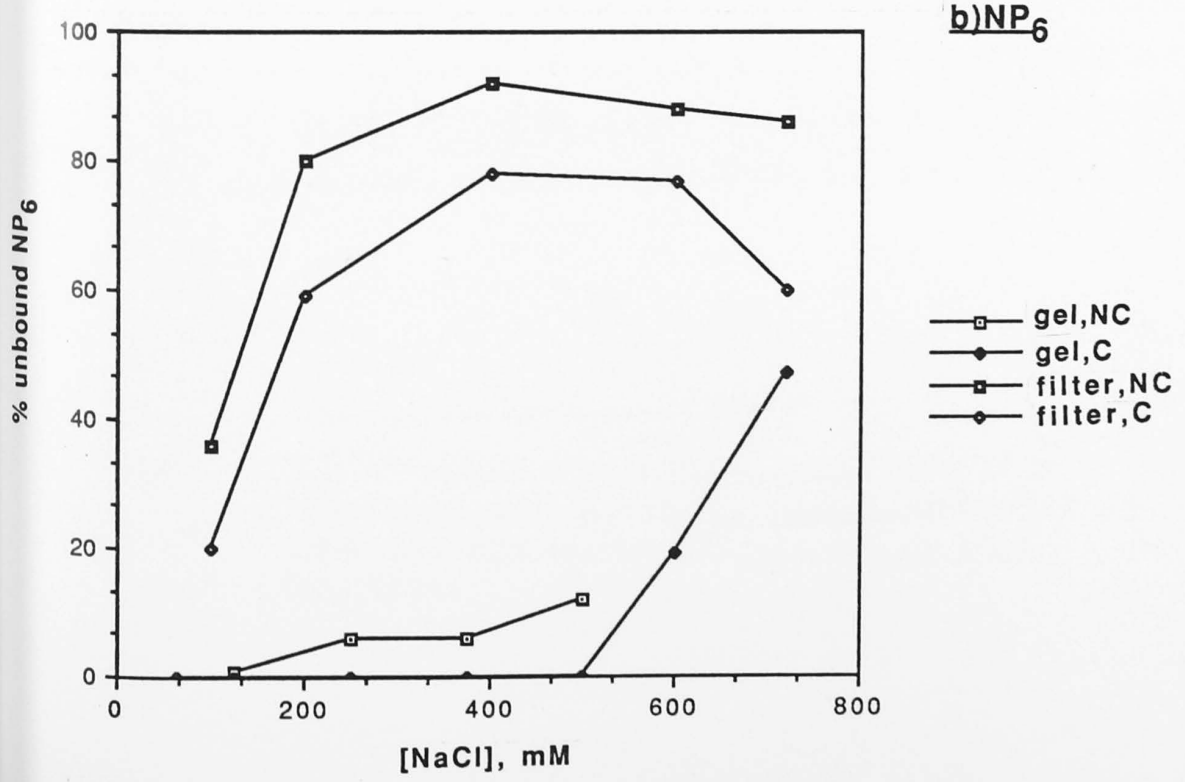


Figure 4.10: Autoradiogram showing electrophoretic retardation with increasing *rough* homeodomain concentrations of the NP_6 and 132 bp DNA fragments

Binding reactions: [α-^{32}P]dATP end-labelled DNA fragments (NP_6 and 132 bp) were mixed with varying amounts of *ro*HD in binding buffer (20 mM Tris pH7.5, 75 mM NaCl, 1 mM DTT, 10% glycerol) to give a total reaction volume of 30 µl. Reactions were incubated at room temperature (25°C) for 30 minutes, and then electrophoresed.
Reactions: **NP_6** (0.44 nM), [*ro*HD] = **A**, 0 nM; **B**, 0.56 nM; **C**, 1.1 nM; **D**, 2.8 nM; **E**, 5.6 nM; **F**, 11.2 nM; **G**, 28 nM; **H**, 56 nM; **I**, 140 nM; **J**, 280 nM; **K**, 560 nM; **L**, 1120 nM.
132 bp (1 nM), [*ro*HD] = **A**, 0 nM; **B**, 0.76 nM; **C**, 1.5 nM; **D**, 3.8 nM; **E**, 7.6 nM; **F**, 15 nM; **G**, 38 nM; **H**, 76 nM; **I**, 187 nM; **J**, 374 nM; **K**, 748 nM; **L**, 1500 nM.

Gel electrophoresis: 10% 0.5 x TBE polyacrylamide gel was pre-electrophoresed at 300 volts for 30 minutes at room temperature. After loading the binding reactions, electrophoresis was at 200 volts for 1.5 hours. The gel was then dried and autoradiographed.

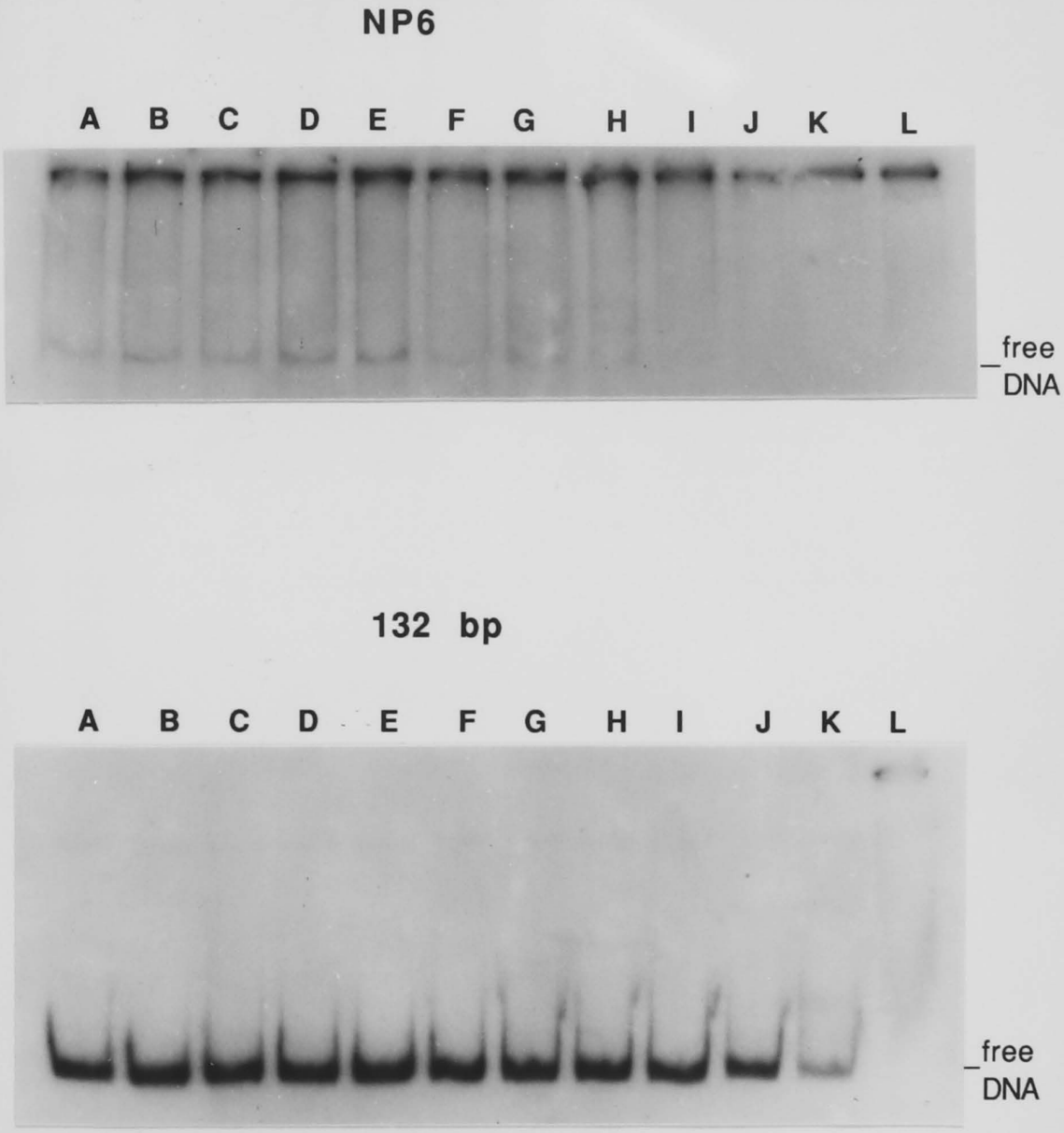


Figure 4.10: Autoradiogram showing electrophoretic retardation with increasing *rough* homeodomain concentrations of the NP_6 and 132 bp DNA fragments.

obtained by filter binding. Thus, for the 67 bp TAA_5 fragment (Figure 4.8a)), the K_d value obtained by gel retardation is around 8.8×10^{-9}M, whilst that obtained by filter binding is ~2.4×10^{-7}M. Similarly, the K_d values for the interaction of the *ro*HD with the 125 bp NP_6 fragment are ~9.8×10^{-9}M and ~1.7×10^{-7}M (Figures 4.8b) and 4.10).

To investigate the affinity of the *ro*HD for DNA fragments around the same size as TAA_5 and NP_6, but possessing a DNA sequence corresponding to the polylinker sites flanking the homeodomain consensus sequences in TAA_5 and NP_6, an 82 bp and a 132 bp fragment were prepared as described in Methods. Binding curves for the interation of the *ro*HD with the 82 bp fragment (Figure 4.8c)) gave K_d values of 1.9×10^{-7}M by gel retardation and $>2 \times 10^{-6}$M by filter binding (since a complete curve was not obtained). Similarly, with the 132 bp fragment (Figure 4.8d)) a figure of 5.9×10^{-7}M was obtained by gel retardation, a $K_d > 2 \times 10^{-6}$M being suggested by filter binding. These results and their ratios to each other are summarised in Table 4.2.

Clearly, the affinity of the *ro*HD for the TAA_5 and NP_6 fragments is stronger than that for the non-specific 82 bp and 132 bp fragments. However, the accuracy of the estimates for the dissociation constants relative to their "true" values is anything but clear. Although, the *ro*HD seems to have approximately equal afffinity for both the TAA_5 and NP_6 fragments as shown by both methods, the 27-fold (TAA_5), and 17-fold (NP_6) discrepancies in the K_d estimates derived by each method is large.

There are a number of explanations which might suggest that the result obtained by gel retardation is lower than in actuality, whilst that obtained by filter binding is somewhat higher.

A common criticism of the filter-binding assay is the loss

of specific protein-DNA binding during the washing of the filter (Riggs et al.1970; Garner and Revzin 1981; Hendrickson 1985). Riggs et al. (1970) found that not all of the DNA bound to protein is retained on the filter. They obtained an average maximum binding equivalent to 30% of the initial DNA concentration, although their best experiment gave a 70% retention. Three reasons for this low plateau of maximum binding were given: first, some DNA fragments may be too damaged to be bound by the protein; secondly, some complexes are lost in washing the unbound DNA from the filters; and finally, some protein-DNA complexes pass straight through the filter when applied. Given the large size of their DNA fragment (~46,000 bp), it is not surprising that damage and loss during washing were found to be the main contributing factors to their low plateau, although the tetrameric *lac* repressor that they studied binds very tightly (K_d ~1 x 10^{-13}M) (Riggs et al.1970). The loss of bound DNA and complexes during washing has been mentioned elsewhere (Fried and Crothers 1981), however, Riggs et al. found that if allowances were made for this "retention efficiency" then it can be corrected for in all calculations. Similarly, the shape of the binding curves and the point of half-saturation were found to be independent of washing. The filter-binding experiments with the *ro*HD gave a range of results, with a plateau of 40 to 100% of the input DNA. The results for the standard binding curves shown in Figure 4.8 suggest a plateau below 100%. However, even allowing for a "retention efficiency" of ~80% in these experiments, the apparent K_d is only lowered by 28% for TAA_5 to 1.7 x 10^{-7}M, and by 18% to 1.4 x 10^{-7}M for NP_6.

Conversely, problems with background can be found with filter binding (Hendrickson 1985). During the experiments with the *ro*HD, upto 11% of the input NP_6, and 6% of the TAA_5 fragments, were retained upon the filter in the absence of the *ro*HD. However,

allowing for such background activity by the filter does not significantly alter the result obtained.

The DNA-protein interaction may be perturbed when the protein binds to the filter (Garner 1986). There have been problems with detecting kinetically labile complexes which dissociate during the washing procedure, for example, *lac* repressor-operator complexes formed in the presence of IPTG could not be detected by filter binding but were evident with gel retardation (Fried and Crothers 1981). Similarly, the specific binding of CAP (catabolite activator protein) to DNA has been demonstrated only at low ionic strength with filter binding, whilst gel retardation has permitted the study of unique CAP-DNA complexes under salt conditions nearer the *in vivo* situation (Garner and Revzin 1981).

Interactions of the *ro*HD with NP_6 and TAA_5 under conditions of increasing [NaCl] (Figures 4.6 and 4.9) again show a tendency for the gel retardation technique to retain higher levels of protein-DNA complexes than are evident from filter binding experiments under the same conditions. Thus, the loss of specific binding with increasing ionic strength is shown to occur quite sharply by filter binding above 100mM NaCl, but does not become evident until higher [NaCl] with gel retardation (above 400mM NaCl for TAA_5, and even higher for NP_6).

The *lac* repressor-operator DNA complex has been shown to be subject to irreversible destruction as a consequence of vigorous mixing, presumably due to shear forces either in solution or acting at a surface (Fried and Crothers 1981). Fried and Crothers (1981) found it difficult to avoid such shear degradation entirely, even with careful hand mixing and slow pipetting onto the gel. Such a disruption of the protein-DNA complexes would presumably be greater during filter binding, notably at the washing stage, where the rate of flow through the filter and the volume of

the washing buffer, may have such an effect.

On the other hand, the binding affinity values obtained by gel retardation may be lower than the true value due to the stabilising of protein-DNA complexes by the gel matrix, and mixing of the reaction mixture with the electrophoresis buffer during loading of the gel (Garner and Revzin 1986).

It has been noted that as long as the free protein concentration in a protein-DNA band on the gel is low compared to the concentration of the complex (e.g. 1:100), there will be little loss of protein as the band moves through the gel, and hence little complex dissociation (Fried and Crothers 1981). In addition, the volume excluded by the gel is inaccessible to reaction components thus increasing the local concentration of reactants, favouring their association. Fried and Crothers (1981) noted that it was possible that the equilibrium binding constant is increased by interactions with the gel. A "caging" mechanism was proposed, whereby the gel provides a "cage" for the protein-DNA complexes such that dissociated components cannot readily diffuse away, and hence must recombine with the same partners. This was supported by studies of *lac* repressor protein exchange within the gel (Fried and Crothers 1981), but was not an apparent problem for CAP interactions with the *lac* promoter (Garner and Revzin 1986).

Such interactions with the gel matrix may partly explain how many different, and some quite small, proteins, can alter the mobility of a DNA fragment to a significant but unpredictable extent (Hendrickson 1985).

For quantitative measurements of binding, the fraction of DNA-protein complexes in solution must remain unchanged during gel loading as well as during electrophoresis. The problem of dilution of the reaction mixture with the electrophoresis buffer during the "dead" time between pipetting the mixture into the loading well, and the unbound DNA and protein-DNA complexes

entering the gel has been investigated (Fried and Crothers 1981; Garner and Revzin 1981). Fried and Crothers (1981) found that it took around 80 seconds for a 203 bp DNA fragment to enter a 5% polyacrylamide gel in TE (10mM Tris, 1mM EDTA, pH 7.4) at a 10 volts/cm voltage gradient. Garnerand Revzin (1981) found that it took less than 3 minutes for a discrete DNA band to enter a 7.5% gel in TBE buffer at 6 mA, and noted that if the retardation of the DNA in a protein complex relative to the free DNA is all that is sought then, once the band of free, uncomplexed DNA has entered the gel, any dissociation of complexes at or near the top will not affect the results, since the DNA released will trail the free DNA throughout the course of electrophoresis.

Changes in the ionic composition of the reaction mixture may occur due to the layering of the solution onto the gel (Garner and Revzin 1981). Despite the addition of glycerol to the reaction buffer some mixing probably occurs during layering, such that the final ionic condition just before the start of electrophoresis is a mixture of the reaction and electrophoresis buffers. The electrophoresis buffer used in the *ro*HD gel retardation experiments, 0.5 x TBE (45mM Tris, 45mM borate, 1.25mM EDTA, pH 8) is a low ionic buffer, thus the mixing of buffers will result in a solution of lower ionic strength than the reaction buffer (20mM Tris, 1mM DTT, 10% glycerol, 100mM NaCl, pH 7.5). This would also increase the stability of protein-DNA interactions and would be disadvantageous if it led to binding of DNA molecules which would otherwise have been free in solution at the full ionic strength of the reaction buffer (Garner and Revzin 1981).

Thus, the binding curves obtained by gel retardation in Figure 4.8 may be lowered due to a dilution of the reaction buffer. Changes in the stability of protein-DNA interactions in the gel are unlikely to affect the result much since the curve was obtained by quantitating the relative changes in the concentration of the

uncomplexed DNA band.

The results obtained by gel retardation for the changes in binding of the roHD with increasing [NaCl] (Figure 4.9) may be affected by dilution of the reaction mixture with electrophoresis buffer during electrophoresis, since the values given for the [NaCl] may be significantly lower, bringing the curve obtained closer to that obtained by filter binding.

Extra protein-DNA interactions due to dilution of the ionic strength of the reaction buffer may be decreased by the addition of non-specific competitor DNA (Garner and Revzin 1981). Binding curves obtained by both gel retardation and filter binding with increasing [NaCl] and 1µg pCE30 as non-specific competitor DNA (Figure 4.9) suggest that, if anything, the stability of the specific *ro*HD-DNA complexes is enhanced by this measure. The wide discrepancy between the results obtained by gel retardation and filter binding is still clearly evident.

Another way of reducing the dilution of the reaction mixture by the electrophoresis buffer is by placing the reaction buffer on top of the gel before loading the reaction mixtures (Garner and Revzin 1981), although this lengthens the time taken for DNA to enter the gel under electrophoresis, if the reaction buffer is of higher ionic strength than the electrophoresis buffer.

A similar discrepancy of an order of magnitude in the results obtained for dissociation constants by gel retardation and filter binding was observed for the *lac* repressor-operator (Fried and Crothers 1981, Riggs et al.1970). However, in this case, it was the figure obtained from the gel retardation experiment which was higher, i.e. weaker binding (Fried and Crothers 1981). Yet, although the operator sequence was the same for both experiments, the size of the DNA fragments used was different- 46,000 bp for the filter-binding study (Riggs et al.1970) and 203 bp for the gel retardation (Fried and Crothers 1981) - and so, the difference in affinities may

be a function of the size of the DNA fragment (Fried and Crothers 1981).

There are a number of estimates of the dissociation constants for homeodomain protein-DNA interactions with which the results obtained for the *ro*HD may be compared. The *Ultrabithorax* protein, *Ubx*Ib (Beachy et al.1988) was found to exhibit a range of affinities for consensus sequences from its own, and the *Antennapedia*, promoter. A K_d of ~ 10^{-8}M was given for the strongest interaction between *Ubx*Ib and the TAA-rich A-1 region of the *Antp* promoter. This figure was determined by DNAase I footprinting of the region with increasing [*Ubx*]. However, Beachy et al. (1988) noted that the K_d may be higher since the effect of the non-specific competitor poly d(I-C) (polydeoxyribonucleotides and inosine and cytosine) present in the reaction mixture, and the specific activity of the *Ubx*Ib preparation, was not known. Indeed, this figure is at the lower end of binding affinities reported for purified mammalian and *Drosophila* transcription factors which have been found to have dissociation constants from 10^{-9} to 10^{-12}M (Beachy et al.1988). However, the clustering of *Ubx* consensus sites near the promoter has suggested a cooperativity of binding, which may increase the overall affinity of *Ubx* for the whole region (Beachy et al.1988).

Hoey and Levine (1988) do not give an estimate of K_d, but note the relative binding affinities of several full-length homeodomain proteins for a number of consensus binding sites. Thus, *zerknullt* (*zen*) has a 250-fold stronger preference for the k1 as opposed to the k2 site in the *engrailed* promoter, and binding to k1 is 100-fold tighter than to the e4 site from the *even-skipped* (*eve*) promoter. Meanwhile, *eve* binds twice as tightly to the k1 site as the k2, and the *paired* and *engrailed* proteins have around an equal affinity for both k1 and k2. These results were obtained by

DNAase I footprinting.

Muller et al. (1988) give dissociation constants for the binding of the 68 amino acid *Antp* HD to a number of oligonucleotides. Binding sites, BS1 and BS2, were characterised in the *engrailed* promoter by DNAase I footprinting, and 26 bp oligonucleotides corresponding to these sites were synthesized (Muller et al.1988). A standard binding curve for each of these oligonucleotides was generated using gel retardation with a range of *Antp* HD concentrations from 6×10^{-10} to 4×10^{-6} M. From this data, dissociation constants of 1.2×10^{-8} M and $\sim 2.5 \times 10^{-8}$ M were obtained for BS2 and BS1 respectively, whilst a non-specific oligonucleotide of the same size had a K_d of $2 - 4 \times 10^{-7}$ M (Muller et al.1988).

Thus, the estimates of the dissociation constants for the binding of the *ro*HD to TAA_5 and NP_6 obtained by gel retardation are in the region of those obtained for other *Drosophila* homeodomain proteins, whilst those derived from filter-binding are higher ($\sim 2 \times 10^{-7}$M). However, the sequence and size of the DNA fragments used are different and so a direct comparison cannot be made.

Further experiments are need to clarify the correct values of the dissociation constants for the interactions of the *ro*HD with TAA_5 and NP_6. Standard binding curves in which there is a low fixed concentration of *ro*HD but a variable DNA concentration (Riggs et al.1970) may complement those obtained by varying the protein concentration. However, if the gel retardation and filter-binding techniques continue to give such consistently different results, perhaps the use of the more accurate DNAase I footprinting method might give better titration results (Galas and Schmitz 1978).

Competition studies may also be conducted to give information on the kinetics of the *ro*HD-DNA binding. The rate

constant for dissociation can be determined by forming equilibrium protein-DNA complexes and then adding an excess of unlabelled specific, or non-specific, competitor DNA to sequester the protein as it dissociates from the radioactively labelled DNA fragment. The order of the reaction and its rate constant may be determined from the fraction of the original complexes remaining with time (Hendrickson 1985). Such kinetic studies have used either the filter-binding (Riggs et al.1970), or gel retardation techniques (Fried and Crothers 1981). A limitation of such kinetic studies by gel retardation is that the protein-DNA complexes should have "lifetimes" greater than around a minute (Fried and Crothers 1981, Garner and Revzin 1981). Otherwise samples are loaded directly onto a gel and electrophoresed; the sampling operation has been achieved in 15 seconds (Fried and Crothers 1981). In these dissociation experiments, it was the increasing concentration of free DNA which was quantitated, since it gave a more accurate result (Garner and Revzin 1981).

Such a competition experiment could be used to determine the relative strength of binding of the *ro*HD for TAA_5 and NP_6 by addition of an excess of one of these species to equilibrium *ro*HD complexes of the other.

4.2.5 Prospects for a Purification Assay for the *rough* homeodomain.

The desirability of an assay for the *ro*HD, which could be used in order to facilitate the purification of sufficient quantities of the protein for structural studies was discussed in Chapter 3. The studies described in this chapter indicate that the *ro*HD has a sequence-specific DNA-binding activity, recognising the TAA_5 and NP homeodomain consensus sequences. Hence, it should be possible to utilise this functional property of the *ro*HD to develop a purification assay for it.

Such assays based on specific protein-DNA interactions have used filter-binding (Rosenfeld 1986) and gel retardation (Hendrickson 1985, Scheidereit 1988) to detect protein-DNA complexes.

The basic criterion for a successful assay are the use of a specific DNA sequence for the protein, and conditions which minimise interference due to non-specific binding by other proteins in the cell extract (Hendrickson 1985). Either the TAA_5 or NP_6 DNA fragments may be used as specific DNA sequences for the *ro*HD. To minimise non-specific binding, an excess of unlabelled non-specific DNA is usually present in the reaction buffer; the cruder the extract, the greater the excess of non-specific DNA that is required (Hendrickson 1985). Under high salt conditions (>0.3 M NaCl) non-specific binding is also reduced (Hendrickson 1985). In purifying cyclic AMP receptor protein and the arabinose-C protein, Hendrickson (1985) found that a level of protein around 0.1% of the total cellular protein was sufficient for their assay by gel retardation.

High levels of non-specific competitor DNA were used in the specific DNA-binding purification assay for nuclear factor I (NF-1) (Rosenfeld 1986), although the level of non-specific DNA binding activity in each protein fraction was also obtained by performing parallel nitrocellulose fitter-binding assays in order to more accurately correlate DNA-binding activity. A homeodomain protein, the B cell-specific octamer-binding transcription factor OTF-2, was monitored during its purification by a DNA-binding assay using gel retardation (Scheidereit et al.1988).

The data obtained from quantitative gel retardation assays using crude lysates have, under conditions which minimise non-specific binding interactions, been found to be identical to those from assays with purified proteins (Hendrickson 1985).

A purification assay for the *ro*HD based on its specific

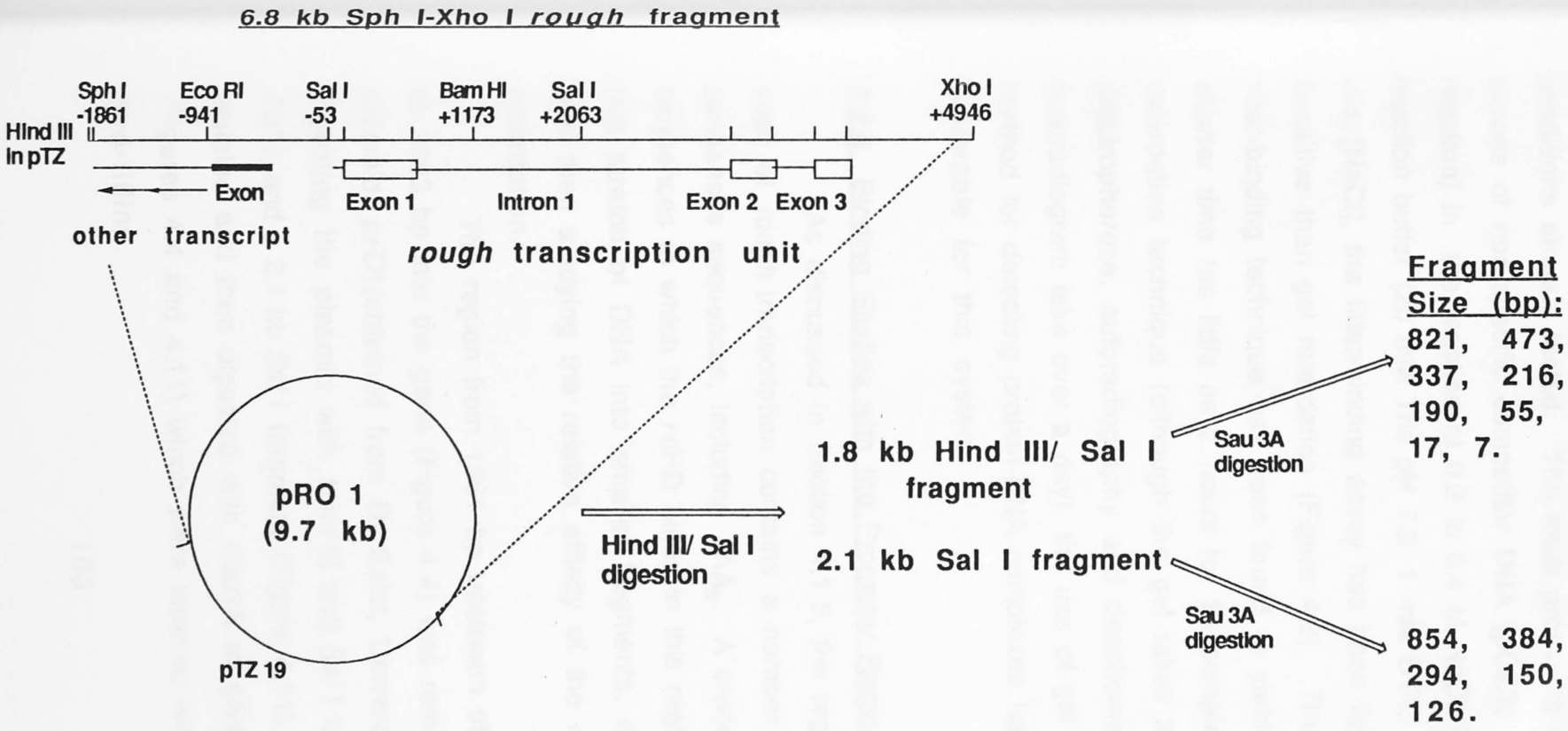


Figure 4.11: The structure of the 6.8 kb Sph I-Xho I *rough* fragment in pRO 1 (obtained from Dr R. Saint) and the DNA fragments generated by restriction enzyme digest of pRO 1 used in the DNA-binding study shown in Figure 4.12.

The sequence of the Sph I-Xho I *rough* fragment, and the sequence composition of the Sau 3A fragments is given in Figure 4.4.

interactions with TAA_5 and NP_6 (at ~10^{-8}M) might start with conditions already tested. The most promising being to include an excess of non-specific competitor DNA (pCE30 at 1-5µg / 40µl reaction) in the reaction at 0.3 to 0.4 M NaCl, in the standard reaction buffer (20 mM Tris pH 7.5, 1 mM DTT, 10% glycerol). At this [NaCl], the filter-binding assay has been found to be less sensitive than gel retardation (Figure 4.9). Thus, although the filter-binding technique has been found to yield a result in a shorter time (as little as 2 hours for 24 samples) than the gel retardation technique (although the gel takes 2-3 hours to electrophorese, autoradiography and densitometry of the autoradiogram take over a day), the use of gel retardation as a method for detecting protein-DNA complexes has been found to be preferable for this system.

4.2.6 Binding Studies with the Promoter Region of *rough*.

As discussed in Section 4.1.5, the region upstream of the start of *rough* transcription contains a number of homeodomain consensus sequences, including TAA_5. A crude way of searching for sequences to which the *ro*HD binds in this region is by digesting this stretch of DNA into smaller fragments, end-labelling them, and then studying the relative affinity of the *ro*HD for them by gel retardation.

The region from 1861 bp upstream of the mRNA start site to 2063 bp into the gene (Figure 4.4) was removed from the plasmid, p*RO*1 (obtained from Dr Saint, University of Adelaide) by digesting the plasmid with *Hind* III and *Sal* I to give a 1.8 kb *Sph* I - *Sal* I, and a 2.1 kb *Sal* I fragment (Figure 4.11). These were isolated and then digested with *Sau*3A to give smaller fragments (Figures 4.4 and 4.11) which were labelled with [α-^{32}P]dATP by end-filling.

Figure 4.12: Autoradiogram showing changes in the elctrophoretic mobility of DNA fragments from *Sau*3A digestion of the 1.8 kb and 2.1 kb DNA from pRO1 in the presence of the *ro*HD and *Ubx* Ib extract.

DNA fragments: pRO1 plasmid DNA digested with *Sal* I and *Hind* III (see Figure 4.11) to give 1.8 and 2.1 kb fragments containing -1861 to 2063 of the *rough* locus (see Figure 4.4). These fragments were digested with *Sau* 3A to generate a range of fragments (see Figure 4.4 for sequence composition of fragments):

1.8 kb: 854 / 384 / 294 / 150 / 126.

2.1 kb: 821 / 473 / 337 / 216 / 190 / 55 / 17 / 7.

Binding reactions: [α-^{32}P]dATP end-labelled *Sau* 3A DNA fragments (5 nM) were mixed with *ro*HD (42 nM) and *Ubx* Ib extract (3 µl) in binding buffer (20 mM Tris pH7.5, 1 mM DTT, 10% glycerol) with, or without, non-specific competitor DNA (1 µg pCE30), and varying concentrations of NaCl, to give a total reaction volume of 40 µl. Reactions were incubated on ice for 30 minutes and then electrophoresed.

Reactions: 1.8 / 2.1 kb (5 nM), *ro*HD (42 nM): **A**, 0 nM *ro*HD; **B**, 60 mM NaCl, 0 µg pCE30; **C**, 60 mM NaCl, 1 µg pCE30; **D**, 480 mM NaCl, 0 µg pCE30; **E**, 480 mM NaCl, 1 µg pCE30; *Ubx* Ib extract (3 µl): **F**, 60 mM NaCl, 0 µg pCE30; **G**, 60 mM NaCl, 1 µg pCE30; **H**, 480 mM NaCl, 0 µg pCE30; **I**,480 mM NaCl, 1 µg pCE30.

Gel electrophoresis: 5% 0.5 x TBE polyacrylamide gel (0.5 mm thick) was pre-electrophoresed at 300 volts for 30 minutes at room temperature. After loading 4 µl of the binding reactions, electrophoresis was at 300 volts for 1 hour, 800 volts for 1 hour. The gel was then dried and autoradiographed.

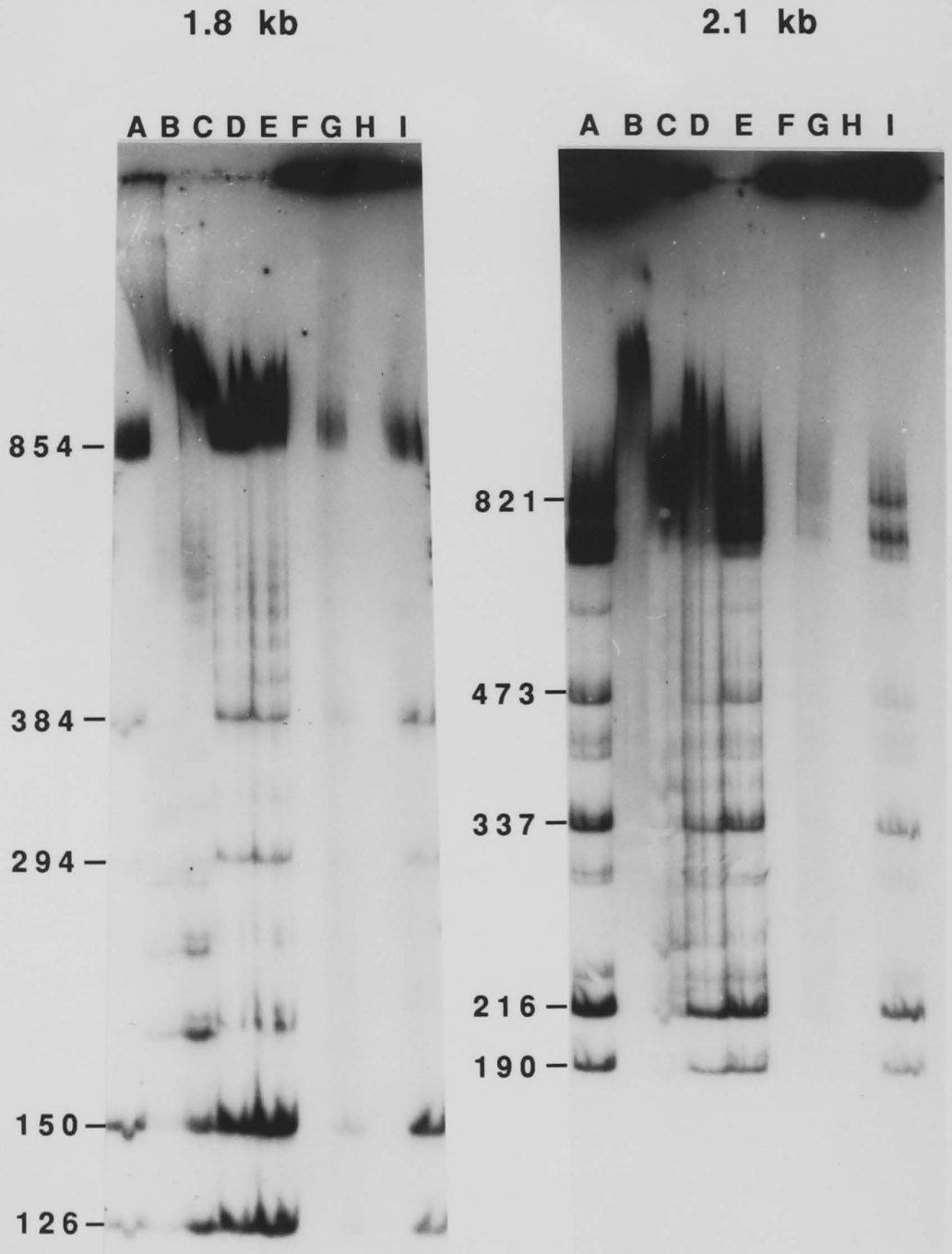


Figure 4.12: Autoradiogram showing changes in the electrophoretic mobility of DNA fragments from *Sau* 3A digestion of the 1.8 kb and 2.1 kb DNA from p*RQ*1 in the presence of the *ro*HD and *Ubx* Ib extract.

Binding of the *ro*HD and *Ubx1b* to these fragments was investigated by gel retardation (Figure 4.12). These studies showed that all of the fragments from both the 1.8 and 2.1 kb pRC1 bands were retarded in their migration through polyacrylamide gels under conditions of low salt (75 mM NaCl) in the absence of non-specific competitor DNA (see Figure 4.12). However under more stringent conditions, some bands seemed to be preferentially bound.

From the 1.8 kb fragment, the 126 and 150 bp bands were less retarded with competitor DNA than the 294, 384, and 854 bp bands, although three novel bands were evident between the 150 and 294 bp bands, even at the most stringent condition tested, (480 mM NaCl + non-specific competitor DNA; lane **E**). There was also a new band above the 294 bp band, and at least six bands between the TAA_5-containing 384 bp band and the largest 854 bp fragment, which was also retarded (Figure 4.12).

The fragments from the 2.1 kb DNA were not retarded to the same extent under stringent conditions (Lane **E**, Figure 4.12). There was no evidence that any of the smaller fragments (7, 17, and 55 bp, not shown), or the largest 821 bp band had been retarded. However, a novel band was evident above the 216 bp band, the lower relative intensity of the 190 bp band indicating that this is probably the one bound; bands were also seen above the 337 and 473 bp bands.

The *Ubx* extract (Figure 4.12) did not seem to bind strongly to any of the fragments under these stringent conditions (Lane **I**), although some ^{32}P was retained in the loading well.

At increasing [*ro*HD] under the same conditions (not shown), higher [*ro*HD] (1000 and 2000 nM) showed a greater relative retardation of all fragments, although the smaller fragments (126 and 150 bp from 1.8 kb; 7, 17, and 55 bp from 2.1 kb) were not as strongly retarded, which suggested that this

additional binding was of a non-specific nature.

Although footprinting experiments of this entire DNA region with the *ro*HD would provide the most detailed information about the sequences recognised by the *ro*HD, there is another procedure which can facilitate the search for binding sites. This entails the isolation of specific protein-DNA complexes using the filter-binding assay, followed by elution of the bound DNA from the nitrocellulose filters; this DNA is then separated by gel electrophoresis, and the relative amounts of the bound fragments may be determined by densitometry (Hennighausen 1987; Beachy et al.1988). Using such a procedure, fragments from the digestion of a 7 kb stretch in the region of the *Antennapedia* gene were screened for *Ubx* binding sites (Beachy et al.1988). The three fragments for which *Ubx* had the highest affinity were subsequently studied by DNAase I footprinting (Beachy et al.1988). Even if a DNA fragment contains a strong binding site, as little as 5-20% of this fragment may be retained on the filter (Hennighausen and Lubon 1987), although Beachy et al.(1988) obtained a 60% retention for their highest affinity fragment. This procedure was not tried for the *ro*HD, but could yield much clearer data than that shown in Figure 4.12 regarding the affinity of the *ro*HD for sites in the DNA studied.

It is difficult to deduce much from these studies, and it is necessary to isolate each of these fragments before testing them for specific binding by the *ro*HD. However, all of the fragments from the 1.8 kb band, which spans the region immediately upstream of the *rough* transcription start site, seemed to be bound under stringent conditions. There also seemed to be some retardation of the fragments from the 2.1 kb band, containing DNA from the transcription start site (190 bp) and the first intron (all but 821 bp, which contains the first *rough* exon). It seems likely that, *in vitro*, the *ro*HD protein can recognise and

bind to sequences in its own gene locus (especially the $(TTA)_5$ sequence). However, this does not necessarily reflect an autoregulatory function *in vivo*.

4.3 Materials and Methods.

4.3.1 DNA Consensus Sequences.

TAA_5 (obtained from C. Desplan (Desplan et al.1988) : The 5'-TAATAATAATAATAA sequence was inserted at the *Sma* I site of the M13mp18 polylinker. An *Eco*RI/*Hind* III digestion of this phage gave just two fragments - one of 67 bp with the TAA_5 sequence; the other is 7.2 kb long.

NP_6 (obtained from C. Desplan (Desplan et al.1988) : The 5'-TCAATTAAATGA sequence occurs 6 times in the orientations: <<.>>>>. This was inserted into the *Bam*HI site of M13mp18. *Eco*RI/*Hind* III digestion gave a 125 bp fragment containing NP_6, and the 7.2 kb fragment.

M13mp18 polylinker sequence (*Eco*RI to*Hind*III) : 5'-GAATTCGAGCTC GGTACCCCGG**GGATCC**TCTAGAGTCGACCTGCAGGCATGCAAGCTT. *Sma*I site is underlined (TAA_5 insertion site), *Bam*HI site in bold (NP_6). Single-stranded M13 preparations of TAA_5 and NP_6 were prepared. Nucleotide sequencing was used to confirm their identity. Double-stranded M13 DNA was prepared by harvesting M13 infected cells (TG1 = [F', *lac*I^q]), using a plasmid preparation method (see Section 2.3.8).

pTZ18u Subcloning and Preparation : The plasmid pTZ18u has the M13mp18 polylinker (Pharmacia). TAA_5 and NP_6 M13 DNA were digested with *Eco*RI and *Hind* III. The DNA fragments bearing the consensus sequences were ligated into the *Eco*RI/*Hind* III digested pTZ18u. TG1 was transformed, and the transformants screened for

the presence of the insert by *Eco*RI/*Hind* III digestion and end-labelling.

Purification of Small NP_6/TAA_5 fragments : The small fragments from *Eco*RI/*Hind* III digestion of NP_6/TAA_5 were purified from a native 5% TBE polyacrylamide gel: the bands were visualised with U.V. following staining with EtBr (0.5mg/litre TBE), and excised. The DNA was eluted from the acrylamide by crushing the bands into elution buffer (0.5 M NH_4OAc, 0.01 M $MgCl_2$, 0.1 mM EDTA, 0.1% SDS) and leaving at 37°C overnight. This mixture was centrifuged, the eluate removed, and the DNA was precipitated by addition of ethanol at -70°C, and by centrifugation at 89,000g for 90 minutes at 4°C.

3102 (obtained from Dr. Saint, Adelaide University) : This is a pBR322 derivative with a 3 kb *Eco*RI fragment from the promoter region of the *Ubx* gene. It includes the consensus sequence TAA_n in the U-A and U-B regions (Beachy et al.1988). Digestion with *Eco*RI/*Hae* III and labelling gave 4 labelled fragments: 273 bp (with the entire U-B region), 162, 34, and 13 bp.

82 bp and 132 bp DNA used in Section 4.2.4 prepared by digestion with *Eco*RI/*Hind*III (82 bp) and *Eco*RI (132 bp), and purification of the fragments from a polyacrylamide gel as with TAA_5 and NP_6. Sequences:82bp:5'-AATTCGAGCTCGGTACCCGGGGATCCTCTAGAGT TCGACGTCACGCGTCCATGGAGATCTCGAGGCCTGCAGGCATGCAAGCT. 132 bp, this is the 82 bp sequence with an additional 40 bp at the 3'end:5'-TGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGA GCTCGAATT.

4.3.2 Protein Preparations.

rough Homeodomain : As discussed in Chapter 3.

Ultrabithorax : The plasmid, pPL *Ubx*-Ib (obtained from L.Gavis, Stanford University; Beachy et al.1988) has the *Ubx*Ib cDNA under

the control of the λ P_L promoter. Its expression to give the 40 kD *Ubx*Ib protein had been tested in a number of strains containing cI857, utilising temperature shift induction, but Beachy et al. (1988) found that the most soluble protein was obtained by using the strain AR120 (a *cI* cryptic lysogen), induction was achieved by addition of nalidixic acid to activate rec A function; rec A then inactivated *cI* repression of P_L. Hence, it was decided to use nalidixic acid induction of AR120 pPL *Ubx*-Ib. This was achieved by addition of 40μg/ml nalidixic acid to the exponential AR120 pPL *Ubx*-Ib culture growing at 37°C, harvesting cells 4 hours later. There was no discernible difference between induced and uninduced extracts on SDS-PAGE so N4830 (carrying cI857) was used as host strain. Induced extracts of this clone showed little improvement on SDS-PAGE gels. Since Beachy et al. (1988) reported that cleared lysates of the AR120 clone were adequate for filter binding, it was decided to use this strain, but to take the *Ubx* protein to a greater purity by following their purification procedure. Thus, induced extracts of AR120 pPL*Ubx*-Ib were prepared by freeze/thaw/ lysozyme lysis and centrifugation at 80,000g for 30 minutes. To the supernatant was added 0.5% Polymin P. After mixing for 20 minutes, this was spun at 7,500g for 15 minutes. This supernatant was brought to 30% ammonium sulphate saturation (as oppose to the reported 20% which had apparently resulted in a loss of 80% of the *Ubx* present), mixed for 1 hour, and precipitated at 45,000g for 30 minutes. This pellet was resuspended in the *Ubx* buffer (50 mM Tris, 5% glycerol, 1 mM DTT, 0.1 M NaCl,pH 7.4) and dialysed against this buffer several times. Since at this stage, there was little encouraging sign on SDS-PAGE gels of a significant enrichment of the *Ubx* protein, it was decided not to embark upon column chromatography but rather to use this extract in the binding studies presented here. 750ml of induced cell culture was

reduced to a volume of 10ml in this final extract which contained a mixture of proteins larger than 20 kdal as visualised on SDS-PAGE gels.

HU I/II: This pure protein solution was prepared during the roHD purification described in Section 3.2.2. Its concentration was estimated as 250µg/ml.

Egg-White Lysozyme (Sigma) : A 10mg/ml solution of the anhydrous protein (Mwt. 14,300) was made up in 75 mM NaCl binding buffer.

4.3.3 South-Western Blot

This technique was adapted from Hubscher (1987). A 7% SDS-PAGE of the protein extracts was run and blotted as for a Western blot (Section 5.3). The nitrocellulose membrane was rinsed twice in TBS, blocked for two hours, rinsed in TBS and then incubated in endlabelled solutions of NP_6/TAA_5 and 3102 at 1µg/ml in 1% gelatin/TBS for 3 hours. The membrane was finally washed in TBS for 30 minutes, air dried and autoradiographed.

4.3.4 Gel Retardation.

The gel retardation studies with the *rough* homeodomain used 40µl reaction volumes in binding buffer (20 mM Tris pH7.5, 1 mM DTT, 10% glycerol, 75-720 mM NaCl). DNA concentrations were assessed using A_{260} with the extinction coefficient of 13,000 $M^{-1}cm^{-1}$ per mole of base pairs. Binding reactions were at temperatures indicated (0°C, ~25°C, 30°C) for the times indicated (3o minutes or 1 hour). The DNA and protein-DNA complexes were separated on polyacrylamide gels (5, 10, or 15% as indicated) with 0.5 x TBE buffer (45mM Tris, 45mM borate, 1.25mM EDTA, pH 8.0). Gels were pre-run at 300 volts for at least 30 minutes. After loading binding reactions, gels were electrophoresed for a further

2 hours at 300 volts (or as indicated). Gels were electrophoresed at 4°C or ~25°C as indicated. They measured 1.5 x 200 x 150 mm (except those shown in Figure 4.12 = 0.5 x 200 x 400 mm). Gels were soaked in 5% glycerol, 0.5 x TBE for 30 minutes, dried and autoradiographed.

4.3.5 Preparation of the TAA_5 Oligonucleotide.

The 15-mer single-stranded oligonucleotides, 5'-TAA_5 and 5'-TTA_5 were synthesized using an Applied Biosystems 380B DNA synthesizer by staff at the John Curtin School of Medical Research. These were presented in ammonia solution, rotary evaporated to dryness, and resuspended in TE. The concentration of the nucleotides was determined from the A_{260}, using an extinction coefficient of 38,500 for TAA_5 (Mwt. 4905), and of 74,800 for TTA_5 (Mwt. 4860). The TAA_5 oligonucleotide was labelled by phosphorylating it with [γ-^{32}P]dATP (10mCi/ml; Amersham) and polynucleotide kinase (Boehringer Mannheim) in kinase buffer (50 mM Tris pH7.6, 10 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine, 1 mM EDTA) at 37°C for 30 minutes. This reaction was chased with 0.5 mM ATP for another 30 minutes, and stopped by addition of 100 mM EDTA. The labelled oligonucleotide was purified from unincorporated label by loading it onto a Mono-Q anion exchange column (Pharmacia) in 50 mM Tris 7.5, 50 mM NaCl, running a 50 mM to 1 M NaCl gradient (15 ml). The radioactivity of elution fractions was assessed in a scintillation counter. The concentration of the labelled TAA_5 was determined, and the double-stranded oligonucleotide made by addition of a 10% excess of unlabelled TTA_5 at 60°C and slow cooling. The mismatched oligonucleotides, which were generated to give >15 bp fragments, were made by mixing labelled TAA_5 with a 10% excess of TTA_5 at 65°C before rapidly cooling the mixture on ice. This was then

labelled with [α-^{32}P]dATP by end-filling.

4.3.6 Filter Binding.

This used Millipore nitrocellulose filters (0.45mM pore; 25mm diameter). The binding reactions were identical to those used in gel retardation. These were applied to the filter under suction, and washed with 3 ml of binding buffer. Filters were placed under infra-red illumination until dry. These were then placed in a plastic scintillation vial with 5ml of scintillant (2.5 l toluene, 12.5g PPO, 0.25g POPOB) and counted in the ^{32}P lane of a scintillation counter. The 100% count for the input labelled DNA in each experiment was determined by filtering the input labelled DNA (in 100µl DNA stop solution (10mM EDTA, 0.2g/ml herring sperm DNA) and 1ml NaPPi (10% TCA, 100mM sodium pyrophosphate) through a glass fibre filter after standing for 20 minutes on ice. DNA was retained quantitatively on the glass filter whilst unincorporated ^{32}P dATP is flushed through during washing with 3 ml NaPPi. The dried filters were counted as above.

4.3.7 Quantification of Autoradiograms by Laser Densitometry.

A quatitative evaluation of the autoradiograms was made using as LKB UltroScan XL Laser Densitometer. Multiple line scans were taken over an X-width of 8,000 microns. The scans were processed by the machine's digital integration facility to give the areas of each peak from which the relative quantities of DNA in the unbound DNA bands could be determined. Care was taken not to exceed the upper limit of the detection system (4 AU) by not using over-exposed autoradiograms. To assure linearity of the results, intensifier screens were not used during autoradiography.

CHAPTER 5

Immunochemical Studies of the *rough* Protein.

5.1 Introduction.

There have been many accounts of the use of immunohistochemical studies of the spatial and temporal location of *Drosophila* proteins *in vivo*. Gene products studied in this way include the homeodomain proteins, *engrailed* (DiNardo et al.1985), *fushi tarazu* (Carroll and Scott 1985), *Ultrabithorax* (White and Wilcox 1984), and *sevenless* (Banerjee et al.1987), which is another of the proteins involved in eye ommatidial development (Section 1.2).

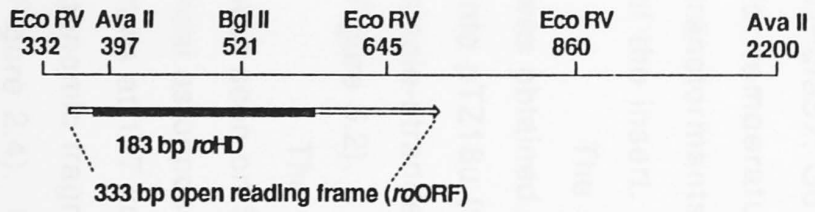
All of these studies entailed the construction of hybrid genes with all, or part of, the coding region of the genes. These hybrid genes were expressed to give fusion proteins against which polyclonal antibodies were raised.

It was decided to follow a similar procedure to study *rough* immunochemically.

5.2 Results and Discussion.

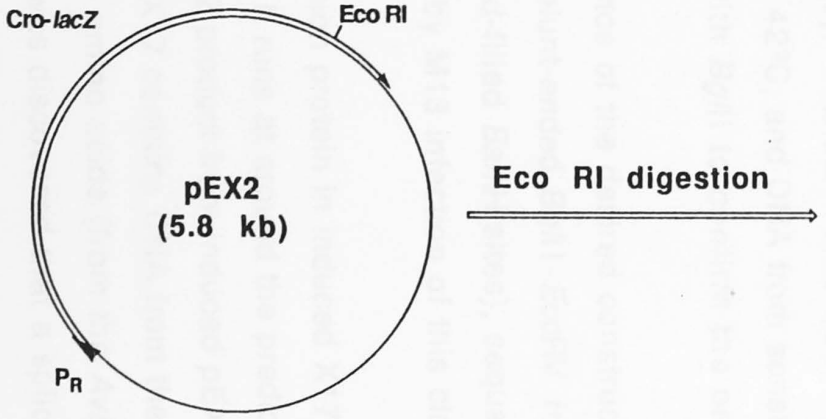
5.2.1 Preparation of the Fusion Gene.

The intention here was to produce a hybrid gene encoding a fusion protein of β-galactosidase/*rough* HD by fusion of the *rough* homeobox to the 3' end of the *E. coli lacZ* gene. The vector system used was pEX, a series of bacterial expression vectors derived from a *Cro-lacZ* gene fusion plasmid which expresses large quantities of fusion protein under the control of the P_R promoter of bacteriophage lambda (Stanley and Luzio 1984). A polylinker present at the 3' end of the *lacZ* gene permits insertion of foreign coding DNA in all three translational reading frames.

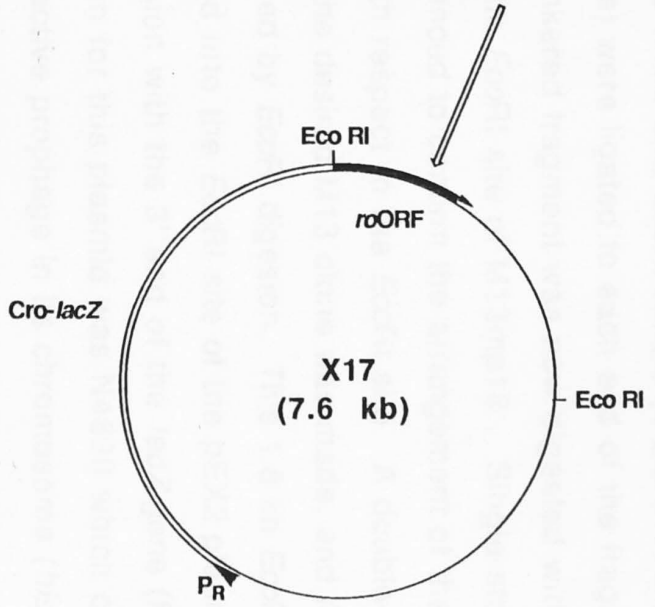


DNA fragment containing the *rough* homeobox; sequence from 332-863 is given in Figure 2.4a).

i) Ava II digestion;
ii) end-fill 1.8 kb fragment;
iii) ligate Eco RI linker (CGGAATTCCG)
iv) Eco RI digestion;
v) ligate into M13mp18, confirm construct by nucleotide sequencing;
vi) Eco RI digestion;
vii) ligate 1.8 kb fragment into pEX2.



Induction of pEX2 in a cI857 *E.coli* strain gives a 116.5 kdal Cro-*lacz* protein.



Induction of X17 in a cI857 *E.coli* strain gives a 125 kdal Cro-*lacz*-*ro*ORF fusion protein (see Figure 5.3a).

Figure 5.1: Construction of the plasmid, X17. The correct insertion of the *ro*ORF was confirmed by nucleotide sequencing (see Figure 5.2)

The *rough* homeobox sequence to be inserted was removed from the 3 kb *Eco*RI-*Sal* I *rough* genomic fragment (Figures 1.1 and 2.4). This was digested with *Ava*II to give a 1.8 kb fragment (Figure 5.1). To prepare this for insertion into the pEX plasmid, the *Ava*II sites were end-filled and *Eco*RI linkers (CGGAATTCCG from New England Biolabs) were ligated to each end of the fragment (Figure 5.1). The linkered fragment was now digested with *Eco*RI and subcloned into the *Eco*RI site of M13mp18 . Single-stranded M13 DNA was sequenced to confirm the arrangement of the 5' end of the homeobox with respect to the *Eco*RI site. A double-stranded DNA preparation of the desired M13 clone was made, and the *rough* fragment was removed by *Eco*RI digesion. This 1.8 kb *Eco*RI fragment was ligated into the *Eco*RI site of the pEX2 plasmid to give an in-frame fusion with the 3' end of the *lacZ* gene (Figure 5.1). The host strain for this plasmid was N4830 which carries the cI857 gene in a defective prophage in its chromosome (*his bio* λ^{def}*cI857*; Gottesman et al.1980). Transformants were screened for temperature sensitivity at 42°C, and DNA from sensitive transformants was digested with *Bgl*II to confirm the orientation of the insert.

The nucleotide sequence of the desired construct, X17, was obtained by subcloning a blunt-ended *Bgl*II-*Eco*RV fragment into pTZ18u (the *Sma* I and end-filled *Bam*HIsites), sequencing the single-stranded DNA obtained by M13 infection of this clone (see Figure 5.2).

The presence of a fusion protein in induced X17 extracts was seen on SDS-PAGE gels. It runs at around the predicted 125 kdal as opposed to the *Cro-lacZ* product from induced pEX2 which runs at 117 kdal (Figure 5.3). X17 contains DNA from the *rough* genomic fragment encoding 95 amino acids (from the *Ava*II site in Figure 2.4). Subsequently, it was discovered that a splice site

Figure 5.2: 3' to 5' nucleotide sequence of the X17 plasmid from the *Bgl* II site in the *rough* homeobox to the fusion with the *lac Z* gene.

The sequence was obtained by sub-cloning an end-filled *Bgl* II-*Eco* RV fragment from X17 into pTZ18u (end-filled following *Sma* I-*Bam* HI digestion), obtaining single-stranded DNA by M13 infection of this clone, and sequencing this DNA using the dideoxy method (see Section 2.3.12). The sequence is of the non-coding strand of the X17 DNA, and hence, runs 3' to 5' from bottom to top for the protein-coding strand. The highlighted regions are: 1, CTTAAG (=GAATTC), *Eco* RI site marking the in-frame fusion of the *lac Z* gene with the *rough* open reading frame (see Figure 5.1); 2, GCTGG (=CGACC), linker ligation to the end-filled *Ava* II site; 3,4, TCTAGG (=AGATCC), ligation of the end-filled *Bgl* II site in the *rough* homeobox to the end-filled *Bam* HI site in pTZ18u.

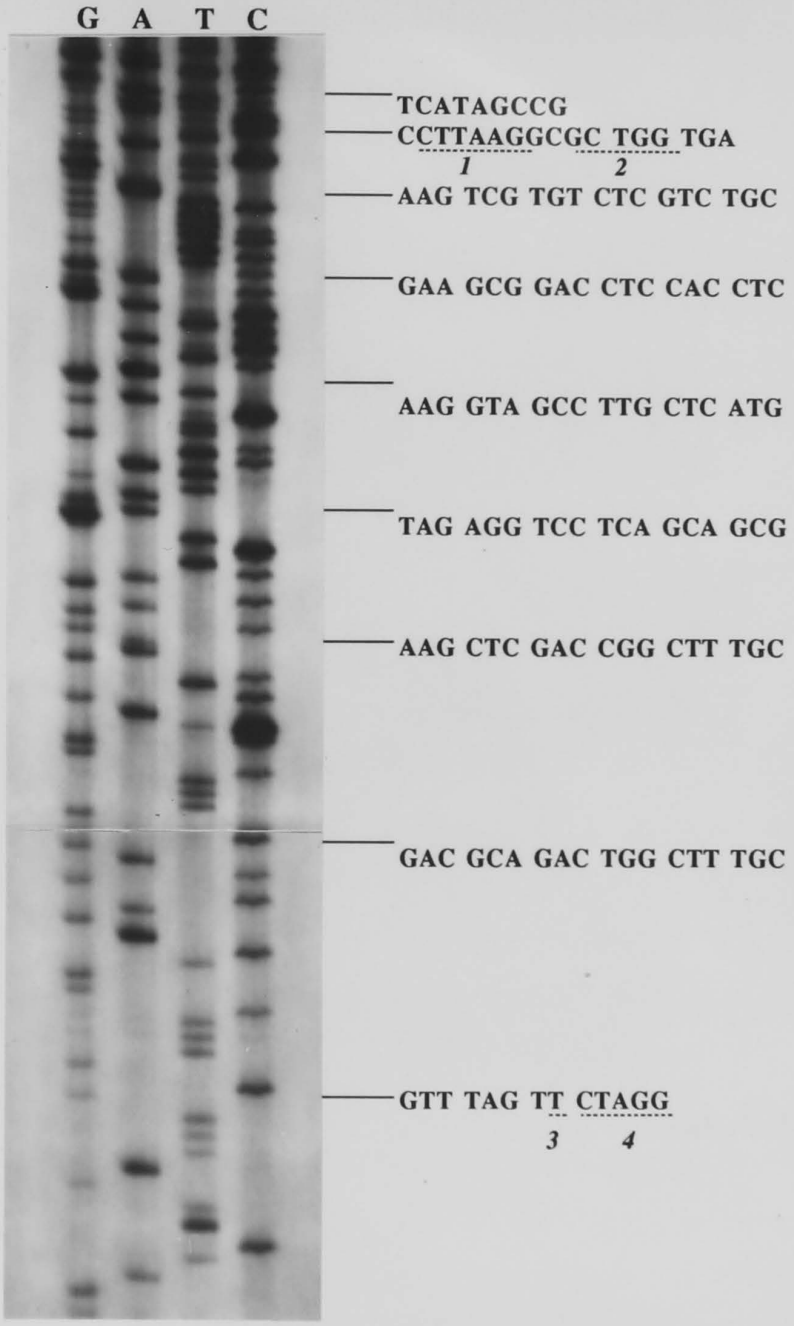


Figure 5.2: 3' to 5' nucleotide sequence of the X17 plasmid from the *Bgl* II site in the *rough* homeobox to the fusion with the *lac Z* gene.

occurs shortly after the homeobox in this open reading frame (Figure 1.1; Tomlinson et al.1988). Thus, the X17 construct encodes 65 amino acids corresponding to the middle exon of *rough* starting 6 amino acids inside the homeobox. 30 superfluous amino acids are encoded in the DNA between the splice site and the stop codon of the open reading frame (Figure 2.4). It was hoped that, although only 65 out of the 1075 amino acids comprising the X17 fusion protein matched the *rough* protein (i.e. 6%), there would be sufficient antigenic sites to generate antisera that would have sufficient activity for use in immunological detection of the *rough* protein.

5.2.2 Preparation of the X17 Fusion Protein.

The X17 fusion protein was produced by temperature shift induction (30 to 42°C) of exponential-phase cultures of X17. Time couse studies suggested that 2 to 3 hours at 42°C gave the best yield of fusion protein (Figure 5.3). The fusion protein is very insoluble unlike the β-galactosidase expressed from pEX2. This is presumed to be due to aggregation (Stanley 1983; Section 3.1). Sonication was necessary to break up the induced cells sufficiently to solubilise the fusion protein in SDS-lysis buffer in order to visualise it on SDS-PAGE gels. To obtain large enough quantities of the fusion protein to use as an antigen, a one litre culture was adequate since the fusion protein accounts for at least 10% of total cell protein. A comparison of lysis procedures showed that sonication was preferable to freeze/thaw lysis in solubilising more of the unwanted bacterial proteins in the lysate. Centrifugation gave a pellet in which the fusion protein accounted for the majority of the protein. To purify the fusion protein from the pellet preparative SDS-PAGE was used, wherein the protein was recovered from the excised gel slice corresponding to the

Figure 5.3: a) SDS-PAGE showing the time course of induction of the fusion protein , X17; b) Western blot showing the binding of X17 antibodies to fractions in a).

a) X17 cells were grown at 30°C to A_{595nm} ~0.5; they were then grown at 42°C to induce expression of the X17 fusion protein. 10 ml aliquots of the induced cell culture were taken at the times shown and the cells were lysed by sonication at 180 Watts for 1.5 minutes. The lysates were centrifuged at 12000g for 10 minutes at 4°C. The pellets were then resuspended in SDS-buffer to A_{595nm}= 10, and heated at 95°C for 10 minutes; 25 µl was loaded onto a 7% SDS-polyacrylamide gel, and separated by electrophoresis. The X17 fusion protein, with a M_r ~ 125 kdal, is indicated.

b) A protein gel similar to that in a) was blotted with nitrocellulose membrane as described in Section 5.3.4. The transfer buffer was 20 mM Tris, 150 mM glycine, 20% methanol, pH 8.3. Transfer was for 15 hours at 25 volts at room temperature. The antibody-staining procedure was as described in Section 5.3.4; the X17 antisera was diluted 1 in 1000; the secondary antibody was peroxidase-conjugated anti-rabbit IgG. The position of the purified X17 fusion protein (3 µg was loaded onto SDS-PAGE), and β-galactosidase (Sigma; 3 µg loaded onto SDS-PAGE), is indicated.

Time of induction (hours)

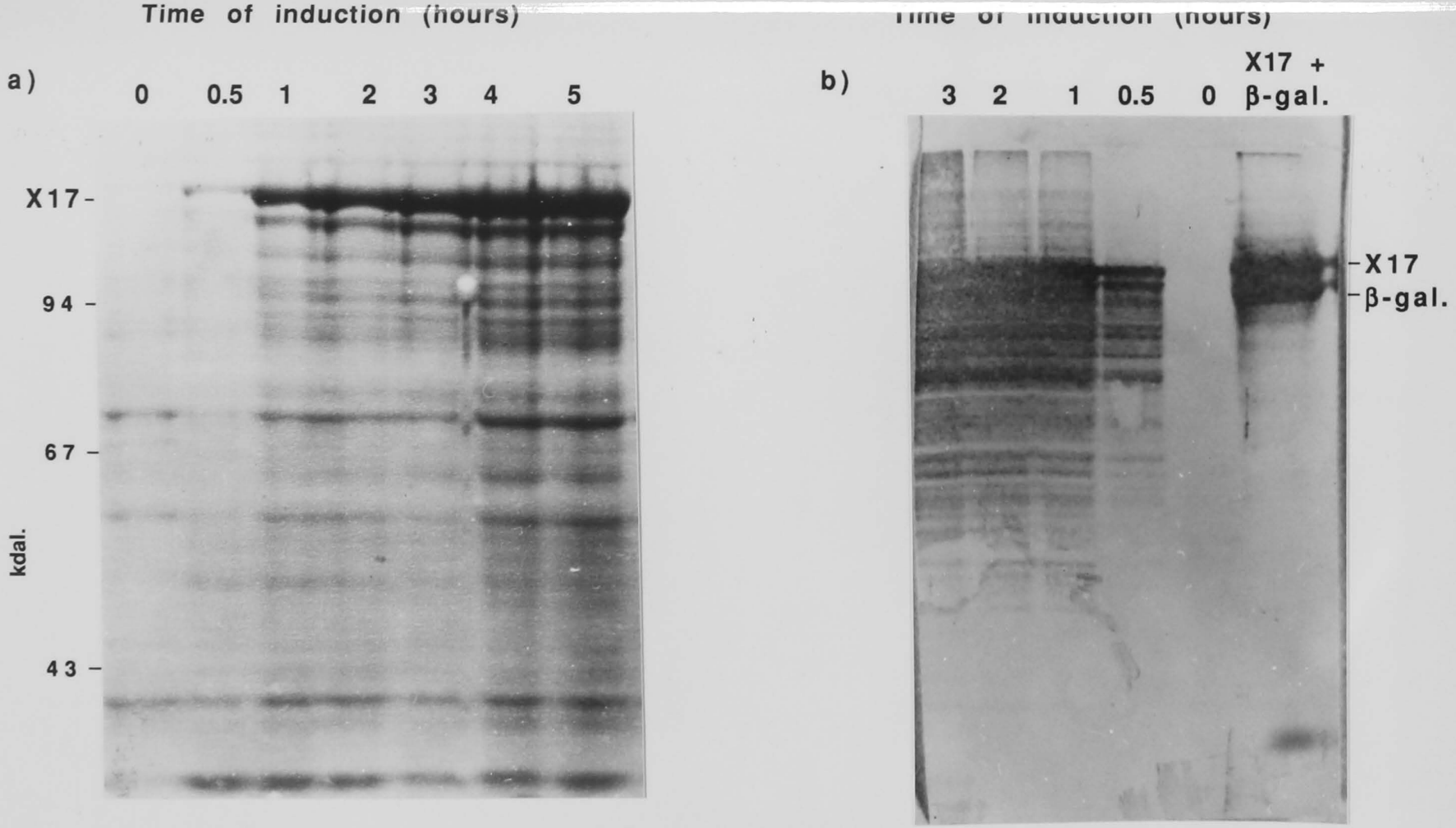


Figure 5.3: a) SDS-PAGE showing the time course of induction of the fusion protein, X17 b) Western blot showing the binding of X17 antibodies to fractions in a).

protein by electroelution (DiNardo et al.1985). The concentration of the recovered fusion proteins was assessed by SDS-PAGE with a range of known concentrations of β-galactosidase. Around 1 mg of X17 fusion protein was purified in this way from 200ml of cell culture.

5.2.3 Raising of Polyclonal Antibodies.

Polyclonal antibodies were raised in two rabbits by injecting them intramuscularly with a 1:1 emulsion of purified X17 protein (130μg) with Freund's Complete Adjuvant (Carroll and Scott 1985). Ten days after the primary injection the rabbits were given a booster injection (150μg X17 in Freunds's Incomplete Adjuvant). The rabbits were bled 10 days later, and given a further booster injection (100μg), before being bled again 10 days later. There were further boosts at 10 day intervals, before the rabbits were bled out.

The specificity of the antibodies for the fusion protein was investigated by Western blotting of SDS-PAGE gels with purified X17 protein, β-galactosidase and crude extracts of induced and uninduced X17 cells (Figure 5.3). Control sera obtained from the rabbits prior to the primary injection did not stain the X17 or β-galactosidase proteins, and relatively little background was evident (not shown). Bleeds taken subsequent to the first booster injection exhibited a strong affinity for the fusion protein. The induced extract showed a wide range of stained protein bands running below X17 and the bacterial β-galactosidase (Figure 5.3). However, the absence of these bands in the uninduced extract suggests that they are protein fragments produced by degradation of the fusion protein, which has been observed with other β-galactosidase fusion proteins (Stanley 1983).

Figure 5.4: a) Western blot showing the binding of X17 antibodies to wild-type (Canton-S) *Drosophila* protein extracts ; b) Western blot showing the binding of X17 antibodies to *Drosophila* wild-type, *ro*^{1}, and *ro*^{w} protein extracts.

a) Protein extracts: **A**, pupal: 30 white (early) pupae were suspended in 200 µl SDS-buffer, and heated at 95°C for 1 hour, 40 µl was loaded; **B**, 100 imaginal eye discs from 3rd instar larvae were suspended in 40 µl SDS-buffer, heated 95°C 1 hour, 40 µl loaded; **C**, 50 3rd instar larvae in 200 µl SDS-buffer as above, 40 µl loaded; **D**, 10 2nd instar larvae suspended in 50 µl SDS-buffer as above, 50 µl loaded; **E**, purified X17 fusion protein (3 µg was loaded onto SDS-PAGE), and β-galactosidase (Sigma; 3 µg loaded onto SDS-PAGE).
SDS-PAGE: 10% separating gel.
Western blot: gel blotted with nitrocellulose membrane as described in Section 5.3.4. The transfer buffer was 20 mM Tris, 150 mM glycine, 20% methanol, pH 8.3. Transfer was at 25 volts for 15 hours at room temperature. The antibody-staining procedure was as described in Section 5.3.4; the X17 antisera was diluted 1 in 500; the secondary antibody was peroxidase-conjugated anti-rabbit IgG. The positions of the two protein bands stained in lane **B** are indicated.
b) Protein extracts: **F**, pre-stained molecular weight standards (BioRad); **G**, 110 Canton-S (wild-type) 3rd instar larvae were homogenised in 300 µl SDS-buffer with a 1 ml glass tissue grinder, and heated at 95°C for 15 minutes, 41 µl was loaded; **H**, 67 *ro*^{1} 3rd instar larvae homogenised in 150 µl SDS-buffer, 45 µl was loaded; **I**, 76 *ro*^{w} 3rd instar larvae homogenised in 200 µl SDS-buffer, 50 µl was loaded.
SDS-PAGE: 7.5 to 15% polyacrylamide gradient separating gel.
Western blot: gel blotted with PVDF membrane as described in Section 5.3.4. The transfer buffer was 20 mM Tris, 150 mM glycine, 20% methanol, pH 8.3. Transfer was at 200 mA for 15 hours at 4°C. The antibody-staining procedure was as described in Section 5.3.4; the X17 antisera was diluted 1 in 500; the secondary antibody was alkaline phosphatase-conjugated anti-rabbit IgG. The positions of the protein bands ~39 kdal in lane **G**, and ~30 kdal in lane **I** are indicated.

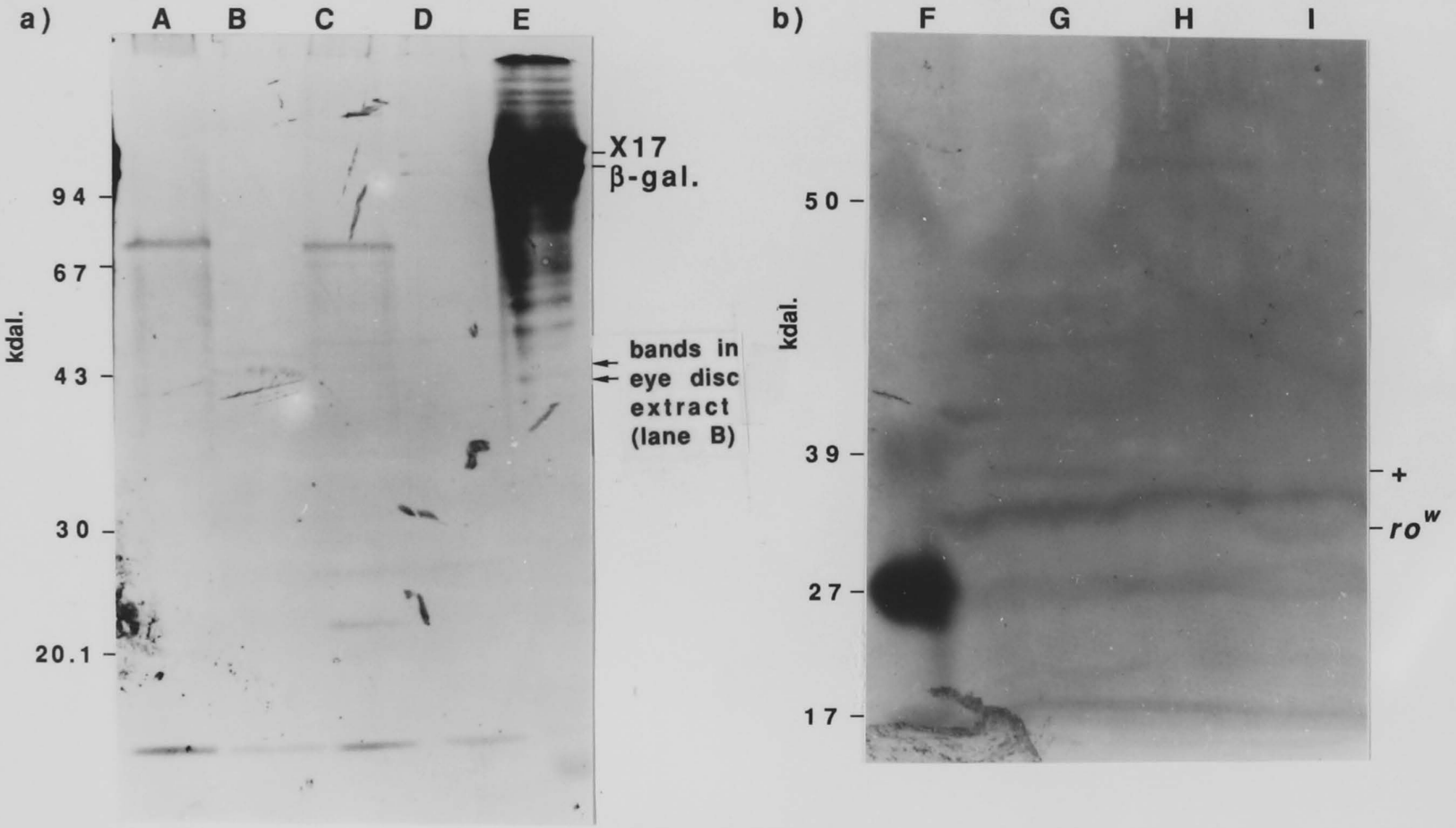


Figure 5.4: a) Western blot showing the binding of X17 antibodies to wild-type *Drosophila* protein extracts; b) Western blot showing the binding of X17 antibodies to *Drosophila* wild-type, ro^1, and ro^w protein extracts.

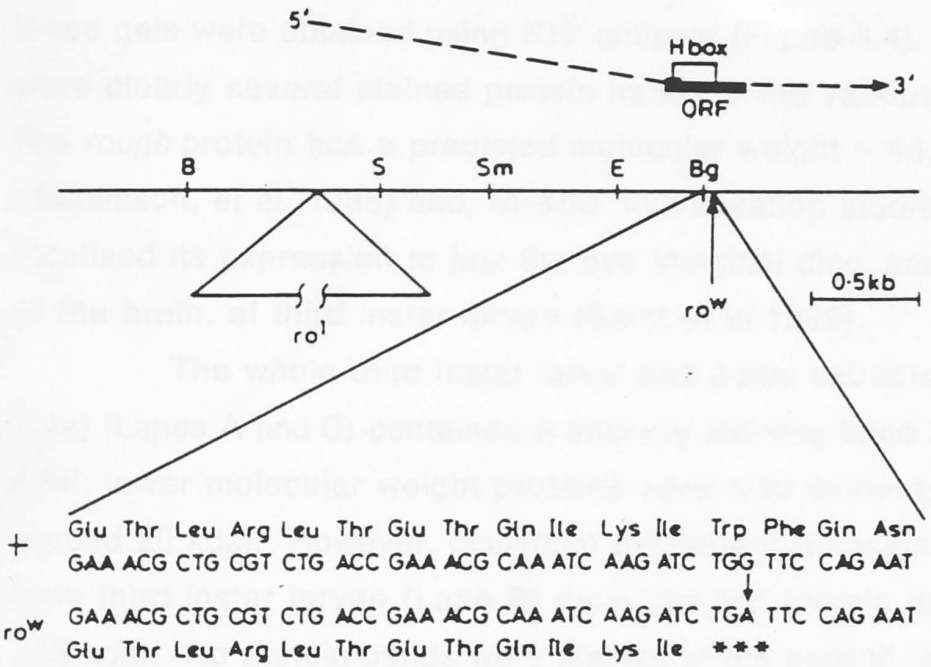


Figure 5.4c): The *rough* mutants, ro^1 and ro^w (from Saint et al. 1988)

ro^1 mutation: a middle-repetitive sequence is inserted between the *Sal*I and *Bam*HI sites, this corresponds to the first intron of *rough* (Tomlinson et al. 1988; Figure 1.1).

ro^w mutation: a single point mutation in the homeobox which changes a Trp codon to a Stop codon. This change is predicted to result in a 237 amino acid protein (wild-type *rough* protein = 350 amino acids).

B- *Bam*HI (+1170) Bg- *Bgl*I (+3549) E- *Eco*RI (+3028) S- *Sal*I (+2063) Sm- *Sma*I (+2582)

5.2.4 Western Blots of Insect Protein Extracts.

Extracts from the larval stages and pupae were prepared in SDS-lysis buffer, and separated on SDS-PAGE. Western blots of these gels were obtained using X17 antisera (Figure 5.4). There were clearly several stained protein bands in the various extracts. The *rough* protein has a predicted molecular weight ~ 43,000 (Tomlinson, et al. 1988) and, *in-situ* hybridisation studies have localised its expression to just the eye imaginal disc, and a region of the brain, of third instar larvae (Saint et al.1988).

The whole third instar larval and pupal extracts in Figure 5.4a) (Lanes **A** and **C**) contained a strongly staining band at ~70 kdal; lower molecular weight proteins were also evident, notably around 20 kdal. However, dissected eye-antennal imaginal discs from third instar larvae (Lane **B**) gave just two protein bands at ~45 kdal. No protein bands were stained in the second instar larvae (Lane **D**). It is possible that one of the bands in the eye disc extract is *rough*.

A Western blot was obtained (Figure 5.4b) with protein extracts of the wild-type Canton-S third instar larvae, and of two *rough* mutants, ro^{w} and ro^{1} (Figure 5.4c); Saint et al.1988). The only significant differences in the staining pattern between the three extracts is the presence of a band migrating at ~38 kdal in the wild-type (Lane **G**) and a protein ~30 kdal in ro^{w} (Lane **I**). This might be explained in terms of the *rough* protein. The 38 kdal band in Lane **G** may be the wild-type *rough* protein, which migrated anomalously with respect to the prestained molecular weight markers due to the highly viscous nature of these extracts. The ro^{1} mutation (Figure 5.4c) is thought to interfere with the expression of *rough* (Saint et al.1988) and, although the coding region of *rough* is unaffected, the levels of *rough* protein expressed in ro^{1} are predicted to be very low relative to the wild-type.

Figure 5.4 d) : Eye-antennal imaginal disc stained with X17 polyclonal rabbit antibodies and a fluorescein-conjugated anti-rabbit secondary antibody (Eric Hines, CSIRO Division of Entomology).

A) (from Tomlinson and Ready 1987) : A late third instar eye-antennal disc; anterior is to the top. The upper portion is the antennal disc which is folded into a series of concentric rings, and below it is the broad, slightly cupped eye disc. The morphogenetic furrow runs dorsoventrically across the eye disc. Ommatidial patterning occurs posterior to the furrow. At the posterior of the eye disc is the optic stalk which carries axons to the brain.

B) A portion of a third instar eye imaginal disc treated with X17 antibodies detected by a fluorescent stain.

Procedure : Eye-antennal discs were dissected from third instar larvae into fixative (4% formaldehyde, 0.2M phosphate pH 7.2), and left to fix overnight at 4°C. Discs washed in TBS (20mM Tris, 0.5M NaCl pH 7.5) overnight at 4°C, then permeabilised with 0.25% Saponin in TBS for 60 minutes at 20°C. Washed in TBS, 0.25% Triton X-100 for 10 minutes, then blocked with 1% BSA for 60 minutes at 20°C. Washed in TBS, 0.25% Triton X-100 for 10 minutes, then incubated in X17 antibodies diluted 1:1 in TBS, 0.25% Triton X-100 for 24 hours at 4°C. Blocked with 1% BSA overnight at 4°C, and washed in TBS, 0.25% Triton X-100 for 10 minutes. Incubated with goat anti-rabbit fluorescein isothiocyanate-conjugated secondary antibodies diluted 1:25 with TBS, 0.25% Triton X-100 for 6 days at 4°C. Washed in TBS, 0.25% Triton X-100 for 2 hours at 4°C, then in TBS for 10 minutes. Whole mount in 90% glycerol in TBS.

The photograph shows the fluorescent light microscopy view of a posterior corner of an eye disc. The morphogenetic furrow runs dorsoventrically (it has not migrated as far as in **A)**). Just behind the furrow a band of fluorescence is evident.

C) : a higher magnification view from the centre of **B)**. The photograph shows the fluorescent band with the morphogenetic furrow above it.

Figure 5.4 d)

A)

MF →

Antennal disc

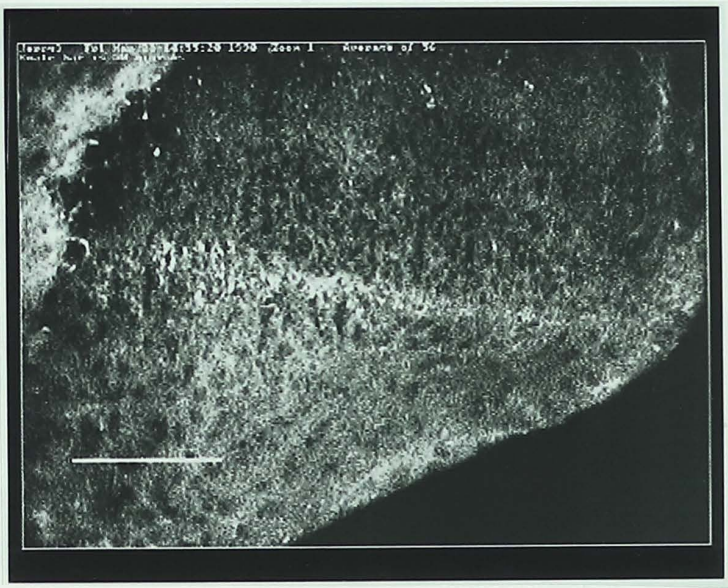
Anterior

Eye disc

Posterior

Optic stalk

B)

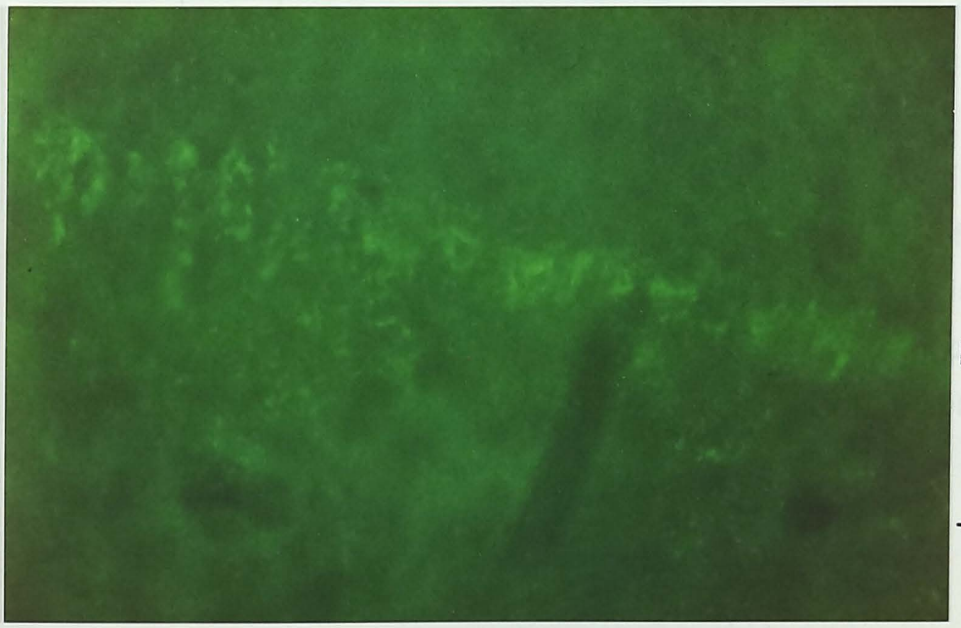


← Morphogenetic furrow (MF)

← Fluorescence

Bar = 50μM

C)



← MF

Fluorescence

Bar = 10μM

Hence, a *rough* protein band if evident, would be much fainter than the wild-type (Lane **H**). Finally, the ro^{w} mutation (Figure 5.4c) is predicted to result in the expression of a truncated *rough* protein (237 amino acids, ~29 kdal). Thus, the band in ro^{w} (Lane **I**) may correspond to this mutant protein.

5.2.5 Immunohistochemistry of Eye-immaginal discs.

Expression of the *rough* gene has been shown at low levels in the developing eye-antennal imaginal disc and a small region of the brain in the third instar larvae (Saint et al.1988). The expression in the imaginal disc was limited to that region destined to give rise to the retina and was maximal at the site of the morphogenetic furrow, although hybridisation was also detected apically in more posterior regions through which the morphogenetic furrow had passed. This pattern of expression is similar to that observed for *sevenless* (Banerjee et al.1987) which also has a role in the correct formation of the photoreceptors (Section 1.1)

The morphogenetic furrow is the site of the early pattern forming events in retinal development (Tomlinson and Ready 1987). Cells in advance of the furrow appear undifferentiated, whilst cells through which the furrow has passed are in various stages of differentiation. Differentiation of cells in this tissue is not lineage dependent but appears to involve a form of recruitment of individual cells into pre-clusters of photoreceptor cells (Tomlinson and Ready 1987).

Immunohistochemical studies using the X17 antibodies were performed on whole eye-antennal imaginal discs of wild-type Canton-S third instar larvae by Mr Eric Hines (CSIRO Division of Entomology, Canberra). These revealed (Figure 5.4d) a specific staining pattern in a band just posterior to the morphogenetic

furrow. This observation compares well with the *in-situ* hybridisation studies which also showed the expression to be localised in the region of the morphogenetic furrow. The phenotypic analysis of *rough* mutants by Tomlinson et al. (1988), showed that the *rough* gene is required for the normal formation of the R3 and R4 photoreceptors (Section 1.2). It appears that photoreceptors R8, R2 and R5 begin to differentiate in the vast majority of developing *rough* ommatidia but the developmental progression incorporating R3 and R4, and subsequent steps are not executed properly. The mechanism and point at which R3 and R4 are affected is unclear (Tomlinson et al.1988).

The immunohistochemical studies of the eye disc were not at a sufficiently high resolution to indicate the precise identity of the cells in which *rough* protein was being expressed, although it might be predicted that to be in the R2 and R5 photoreceptors during the recruitment of the R3 and R4 cells. This may be shown by electron microscopy studies using immunogold-conjugated secondary antibodies.

5.2.6 Immunochemical Assays for the *ro*HD.

As described in Sections 3.2.1 and 3.2.3, immunochemical methods using the X17 antibodies were tried as a means of detecting the *ro* HD to facilitate its purification. However, there were problems experienced in the blotting of proteins for Westerns (Figure 5.5). Although the *ro*HD was successfully blotted onto PVDF from a 10% SDS-PAGE gel for protein sequencing (Section 3.2.1), difficulties were found in transferring it from higher % gels required for a satisfactory visualisation of low molecular weight proteins during purification (Figure 5.5). As can be seen in Figure 5.5, the blotting of elution fractions from an ion-exchange column (**a**) onto PDVF membrane (**b**) resulted in an incomplete transfer pattern (**c**). The transfer of small proteins (less than 27 kdal)

Figure 5.5: a) SDS-PAGE of elution fractions from a BioRex ion-exchange column during *ro*HD purification studies; b) Western blot of fractions in a); c) SDS-PAGE after protein blotting for Western in b).

a) pJG(2) protein extract: Induced pJG(2) cells (net weight 23 g/ see Chapter 3) resuspended in 300 ml lysis buffer (Buffer A, 0.6 M NaCl, 10% sucrose, 0.25 mM PMSF), and lysed at 10,000 p.s.i. in a French Press. Lysate centrifuged at 18,000g for 20 minutes at 4°C. 0.4% Polymin P added to supernatant, mixed on ice for 15 minutes, and centrifuged at 14,000g for 15 minutes at 4°C. Supernatant taken to 60% ammonium sulphate saturation, then centrifuged at 18,000g for 20 minutes at 4°C. Pellet resuspended in 30 ml Buffer A + 0.1 M NaCl, and dialysed against Buffer A + 0.1 M NaCl.
BioRex chromatography: BioRex column (17 x 1.5 cm) pre-equlibrated with Buffer A + 0.1 M NaCl, then loaded with above protein extract. Column washed with 200 ml Buffer A + 0.1 M NaCl, then eluted with 200 ml salt gradient (0.1 to 1.1 M NaCl), collecting 5 ml fractions.
SDS-PAGE: Elution Fractions corresponding to two main peaks (peak one= 20 to 26, max. A_{280nm}= 7.7 (fractions 21,22,23)/ peak two= 26 to 32, max. A_{280nm}= 3.0 (fraction 28)) were loaded (20 µl each) onto an 18% gel and electrophoresed as shown.

b) SDS-PAGE: as in **a)**.
Western blot: gel blotted with PVDF membrane as described in Section 5.3.4. The transfer buffer was 10 mM CAPS, 20% methanol, pH 11.0. Transfer was at 200 mA for 15 hours at 4°C. The antibody-staining procedure was as described in Section 5.3.4; the X17 antisera was diluted 1 in 500; the secondary antibody was peroxidase-conjugated anti-rabbit IgG.

c) The SDS-polyacrylamide gel described above, which was blotted to give the result in **b)**, was stained for protein (using the standard Coomassie Blue stain - see Section 3..3.1) after completion of the blotting procedure in **b)**, showing the incomplete pattern of protein transfer.

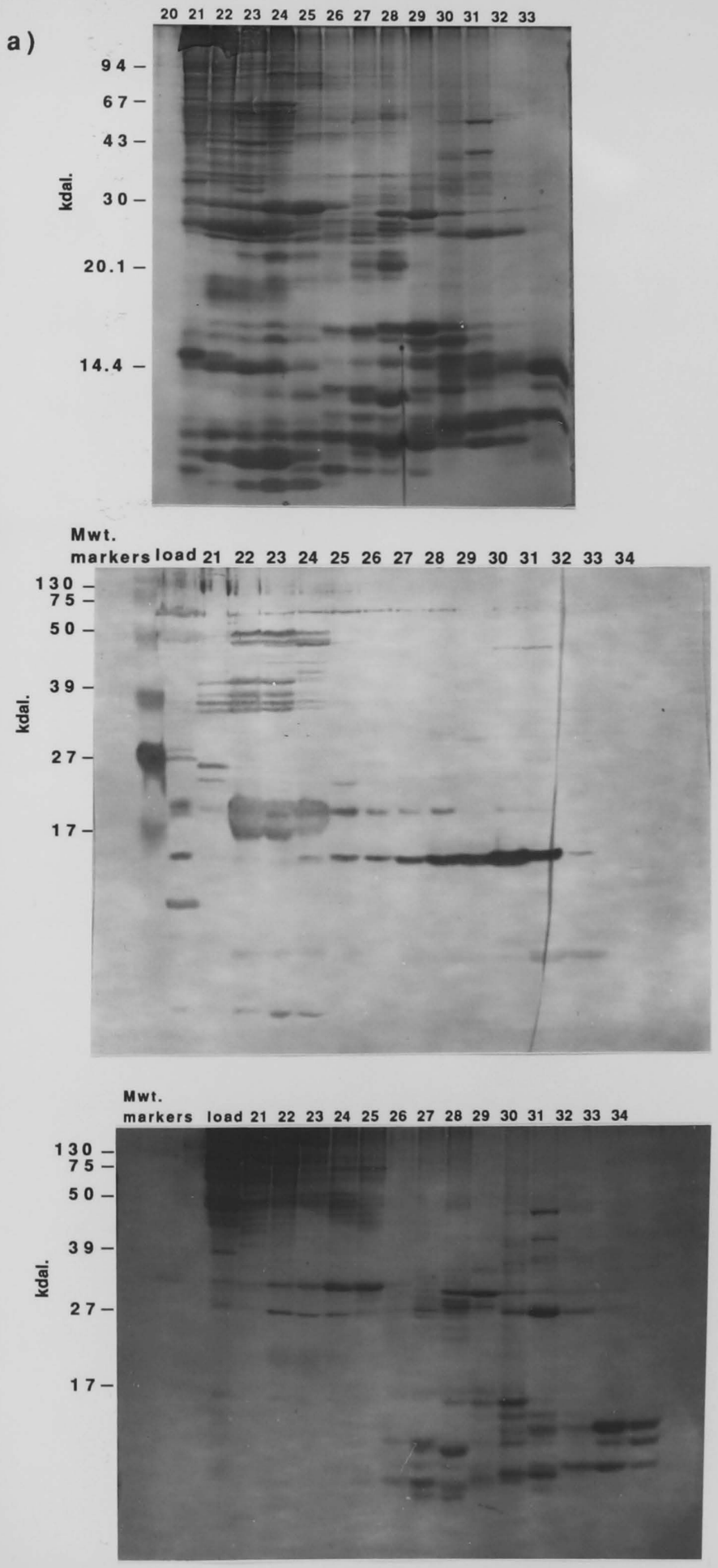


Figure 5.5: a) SDS-PAGE of elution fractions from a BioRex ion-exchange column during *rø*HD purification studies; b) Western blot of fractions in a); c) SDS-PAGE after protein blotting for Western in b).

appears to decrease with increasing basic charge (proteins eluting in later fractions will have a higher positive charge than those eluting earlier). Solutions to this problem regarding the blotting of small basic proteins have been suggested (Szewczyk and Kozloff 1985). These include raising the pH of the transfer buffer to reduce the positive charge on proteins; and minimising the equilibration of the gel in transfer buffer prior to blotting. The methanol in the transfer buffer is required for the retention of proteins on the membrane but removes SDS from the proteins during transfer (increasing the positive charge) (Szewczyk and Kozloff 1985). Transfer buffers at pH 11.0 were tried, and gels were not equilibrated in transfer buffer. However, under these conditions (Figure 5.5), background staining increased, notably of proteins with a similar M_r to the *ro*HD (e.g. the protein ~16 kdal in Figure 5.5a), Fractions 26-32, which stained strongly on the Western blot in 5.5b).

Thus, although transfer is improved at lower % SDS-PAGE (Szewczyk and Kozloff 1985), the lower resolution of small proteins, given the levels of background antibody staining, prevented the development of an immunochemical assay for the *ro*HD.

5.3 Materials and Methods.

5.3.1 Engineering of the X17 hybrid gene.

The methods used in this construction were essentially as presented in the Methods section of Chapter 2.

Preparation of M13 DNA for sequencing (Yanisch-Peron et al.1985) : The ligation mixture of double-stranded M13mp18 DNA with the insert DNA was used to transform competent *E. coli* 7118 (F'

proAB lacI^{q}ZΔM15; Yanisch-Peron et al.1985). The transformed cell mixture was added to 3 ml of top agar containing 20 µl IPTG solution (24 mg/ml), 20 µl X-gal (20 mg/ml dimethyl formamide), and 200 µl log-phase 7118 culture. This was plated on LB agar and grown overnight at 37°C. Blue lysis plaques indicated religated M13 DNA, so clear plaques were selected for DNA preparation. Clear plaques were toothpicked into 2ml of 7118 cells in LB (A_{595}=0.05). This was grown overnight at 37°C. 1 ml was centrifuged and the supernatant (with the M13 phage) removed. Phage was precipitated by addition of 270 µl PEG solution (20% in 2.5 M NaCl), mixing for 15 minutes, and centrifuging. The pellet was resuspended in 200µl TE and the DNA extracted by phenol extraction, with the addition of 100µl TE-saturated phenol, mixing, and centrifugation. The DNA was precipated from the aqueous phase by ethanol precipitation. The single-stranded DNA was resuspended in 20µl TE, 2µl of this solution was adequate for dideoxy sequencing, which proceeded with the annealing of a primer oligonucleotide in a manner similar to that described in Chapter 2. The sequencing protocol followed the recipe with the sequencing kit from BRESA. An easier procedure utilising pTZ18u (Pharmacia) was used later. Here, the ligation mixture was used to transform *E. coli* TG1. These cells were grown on LB plates with X-gal/IPTG top agar. White colonies were selected. Single-stranded pTZ DNA was obtained by adding 5µl of a suspension of M13K07 helper phage (Pharmacia) to a growing culture of the clone, purifying and sequencing it as described above.

5.3.2 Purification of the Fusion Protein.

The induced X17 cell pellet was resuspended in 50 mM Tris pH 7.6, 10% sucrose, and stored at -70C. It was thawed and sonicated at 100 Watts for 4 x 1 minutes on ice using a sonic

probe. The lysate was spun at 10,000g for 30 minutes at 4°C. This pellet was resuspended in protein loading buffer (60 mM Tris 6.8, 4% SDS, 100 mM DTT, 10% glycerol) and boiled for 10 minutes. The equivalent of 100ml of culture was loaded onto a 7% preparative SDS-PAGE gel (3mm x 150mm x 250mm) which was run overnight at 75 volts. The gel was stained in aqueous Coomassie Blue (3% in 4:1 $H_2$0/MeOH) for 15 minutes then destained in 0.25 M KCl at 4°C. The band corresponding to the fusion protein was excised and diced. The protein was eluted from the gel slices by placing them in a BioRad Electroelution apparatus. Elution was overnight at 50mA, the protein was concentrated in the dialysis cups associated with the apparatus (cut-off 20 kdal), and recovered in 0.5ml per cup. The fusion protein concentration was assayed by running against β-galactosidase (from Sigma), made up to 0.65µg/µl in TE.

5.3.3 Raising Antibodies

The protein was emulsified with Freund's Complete or Incomplete Adjuvant from the Commonwealth Serum Laboratories. Injections were made into the hind leg using sterile 24G needles and 1ml syringes. The rabbits were bled by nicking the ear veins. The blood was left to clot overnight at 4°C before centrifuging at 10,000g for 15 minutes. The serum was removed and stored at -70°C . When needed it was thawed and kept at 4°C.

5.3.4 Western Blots (Towbin et al.1979)

SDS-PAGE of proteins were carried out as described in Chapter 3. Blots of these were obtained by placing the gels in a "sandwich" consisting of a Scotchbrite pad, a sheet of 3MM paper, the gel, a sheet of pre-soaked (nitrocellulose membrane (BioRad), or Immobilon PVDF transfer membrane (Millipore)), another sheet of 3MM paper and a Scotchbrite pad. This was assembled in a

BioRad Transblot apparatus, care being taken to remove all air bubbles whilst assembling it in transfer buffer (20mM Tris, 150mM glycine, 20% MeOH pH 8.5). The blot was run overnight at 25 volts, then for 1 hour at 60 volts. Variations in the transfer conditions and buffers used are described in Section 5.2 and Figure legends.

The extent of transfer could be initially assessed from the transfer of the pre-stained protein standards. The membrane was washed in TBS (20mM Tris, 0.5M NaCl pH7.5) for 10 minutes, before blocking any remaining reactive sites with 3% gelatin in TBS for 1 hour with shaking. The membrane was washed in TTBS (0.005% Tween-20 in TBS) and then placed in the primary antibody solution overnight (rabbit antisera diluted in antibody solution=1% gelatin in TTBS). The dilution of rabbit antisera varied from 1 in 500 to 1 in 1000. The membrane was now washed twice in TTBS before being soaked in the secondary antibody solution for 2 hours. The secondary antibodies used were Donkey Anti-Rabbit IgG covalently linked to the Peroxidase enzyme (Amersham), which was used at 1 in 2000 dilution; and goat anti-rabbit IgG alkaline phosphatase conjugate (Sigma) used at 1 in 1000 dilution. The membranes were then washed twice in TTBS, and once in TBS before being developed in the staining solution. The staining solution for peroxidase is 4-chloro-1-naphthol (50mg in 300µl ethanol) in 50mM Tris 7.6 (100ml)/30% hydrogen peroxide (100µl), which gives a purple precipitate. For alkaline phosphatase, the stain is β-Naphthyl Acid Phosphate (25mg), Fast Blue BB salt (25mg) and $MgSO_4$ (60mg) in 50ml boric acid solution (3.7g boric acid, 1.8g NaOH per litre), giving a burgundy red precipitate. The staining reaction is stopped by rinsing in water several times.

CHAPTER 6.

The Secondary and Tertiary Structure of the Homeodomain.

6.1 Introduction

This chapter is concerned with predictions of the secondary and tertiary structure of the *rough* homeodomain. In this respect, extensive reference will be made to the known tertiary structure of the *Antennapedia* homeodomain (Otting et al.1988, Qian et al.1989).

6.1.1 The Folding and Structure of Globular Proteins.

The biosynthesis of proteins during the process of translation takes place on free or membrane-bound ribosomes, yielding distinctive linear sequences of amino acids which, in a "second translation", attain their unique spatial configuration (Jaenicke 1987). Interactions with other proteins or membranes may serve to regulate structure formation in terms of specific subunit recognition, chemical modification, and addressing, finally providing the functional state (Golonbinoff et al.1989).

The spontaneous acquisition of the native three-dimensional structure and the capacity to form higher-order assembly structures are primary functions of the nascent polypeptide chain. The detailed mechanisms of both processes is still not well understood. This applies to co-translational folding steps as well as post-translational processing, addressing, and compartment-alization (Jaenicke 1987).

Studies of the reversible denaturation of globular proteins have shown that folding and association occur in a spontaneous and autonomous way, and that the amino acid sequence contains the necessary information for regaining the native tertiary structure, although directionality of chain elongation *in vivo* may confer

information not available *in vitro*. (Jaenicke 1987).

There is debate as to the question of whether protein folding provides a unique structure for a given amino acid sequence and, if so, does this structure represent the most stable state on the energy profile (Go 1983, Jaenicke 1987). The term "unique structure" is misleading since the protein in its native three-dimensional structure is a dynamic system fluctuating around a limited number of preferred conformations. The corresponding motions involve amino acid side chains, stretches of the polypeptide chain, and domains. Their amplitude and angles differ over a wide range with maximum values of ~10 Å for chain movements and ~20° for domain rotation (Jaenicke 1987).

The "translation" of a given amino acid sequence into its three-dimensional structure may be influenced by co-translational or post-translational processing of the nascent polypeptide chain, as well as the specific solvent environment.

The aqueous or non-polar environment of the nascent polypeptide chain will play a significant role in determining the final protein structure, especially given the ampholytic properties of amino acid residues and the amphipathic nature of structural elements detected in all known protein structures. Proper folding in an aqueous solution implies that interactions between water and the polar and non-polar portions of the protein will play a significant role in determining the shape and stability of the protein because globular proteins minimise their free energy by segregating non-polar amino acids from the aqueous medium (Jaenicke 1987).

The balance of the free energies of the unfolded and folded conformation is found to be a small difference in large numbers: in the denatured state, the chain entropy dominates; upon folding to the native state, the loss in chain entropy is compensated for by

the sum of the intramolecular interactions, and the positive solvation entropy. This latter gives rise to "hydrophobic interactions" as well as a minimum hydrophobic surface area (Jaenicke 1987, Go 1983).

Estimates of the contributions of the different types of molecular interactions to the free energy of protein stabilization show that protein folding and association is determined by the increase in entropy caused by water release from non-polar and polar atoms or residues. Unfavourable free energy terms are derived from the loss of rotational and translational degrees of freedom which invariably accompany intramolecular interactions as well as complex formation. Van der Waals interactions and hydrogen bonds are considered to be less important since these interactions replace similar ones made with solvent molecules in the free unfolded chain. Their main role is to confer specificity in the protein structure (Jaenicke 1987).

In general terms, the structure of proteins is based upon a hierarchical order of protein folding. In an amino acid sequence, next neighbour interactions lead first to regular structural elements of secondary and supersecondary structure, which subsequently merge into domains as structural units that fold independently. Preformed domains, at the next level, coalesce to form the compact tertiary structure (Chothia 1984). In the case of quaternary structure formation, "structured monomers" with native-like tertiary structure acquire their final state of lowest free energy by sequential association and shuffling reactions. (Jaenicke 1987).

6.1.2 Secondary Structure.

In acquiring its three-dimensional structure, it may be assumed that portions of the nascent polypeptide chain can serve as "nucleation sites" for folding because they can "flicker in and

out of the conformation that they can occupy in the final protein" (Jaenicke 1987). Candidates in this regard are α-helices, β-structures, and hydrophobic clusters.

Regular structures in proteins are characterised by specific torsion angles, e.g. an α-helix has phi = 58°, and psi = 47°. The α-helix with 3.6 residues per turn (and other helical structures) is formed by local, next-neighbour interactions. β-structures involve interactions between distant parts of a polypeptide chain, as well as intermolecular interactions of different chains, rather than such next-neighbour interactions within a contiguous sequence of amino acid residues. Studies of high resolution structures show that distortions of α-helices and β-sheets occur frequently due to the tendency of the polypeptide chain to achieve a compact structure with the minimum of empty space. These irregularities are, however, insignificant for the problem of folding and association, because they do not interfere with the relevant features of the secondary structural elements, namely, the dipole nature of α-helices, the amphipathic complementarity and packing of regular arrangements of the polypeptide chain, and their involvement as "seeds" in the initiation of structure formation (Jaenicke 1987).

Reverse β-turns are, after α-helices and β-strands, the next most abundant structural feature in globular proteins. Since they contain mostly hydrophilic residues, they are normally located at the surface of proteins Although β-turns are very important in accomplishing a compact globular structure, they tend to play only a passive role in the folding process. Containing a maximum of one hydrogen-bond, they are a point of least resistance to non-covalent forces tending to bend the polypeptide chain. This assumption seems to be supported by the large variety of observed bend types (> 11), none of which seems to form a particularly stable

conformation.

Other "non-repetitive" structures (commonly, all amino acid residues that are not in α-helices, β-sheets or turns are designated as "random coil") which connect α-helices or β-strands may be either highly organised and well-defined irregular stretches, or genuinely disordered regions caused by protein flexibility.

Apart from the N- and C-terminal stretches of the polypeptide chain, which are often found to exhibit high chain mobility, disordered regions generally have very definite functional roles, e.g. as hinge regions between structural domains, or as specific interaction sites to entrap RNA and DNA in nucleoprotein complexes (e.g. Anderson et al.1987).

6.1.3 Tertiary Structure

The tertiary structure of a protein is basically determined by the localization of regular α- and β-structural segments along the polypeptide chain; by their topological arrangement in space; and by their "assembly" in terms of close packing, i.e. minimization of accessible surface area, generating domains or the compact, solvent-free globular shape.

6.1.4 Domains

Domains, as compact substructures within protein molecules, represent autonomous folding units with a minimum surface area to volume ratio resulting in globules with the most interactions within and the least without. Their size varies considerably with an average of ~150 ± 50 residues.

"Structural domains" show independent folding and independent assembly to a native-like structure. This "independence" refers to *in vitro* folding and translation experiments where nascent multi-domain proteins have been shown to "fold by parts" in a co-translational fashion (Jaenicke 1987). Based on this kind

of vectorial mechanism, with well-defined assembly modules, large polypeptide chains speed up folding by many orders of magnitude; at the same time, they succeed in minimizing incorrect intramolecular long-range interactions.

Generally, domains fold independently in kinetic terms, and they are thermodynamically stable. Folding units that are integrated into a larger entity as subdomains or domains may be considered as obligatory kinetic intermediates on the folding path. Apart from eliminating incorrect aggregation ("wrong aggregation"), they may be of selective advantage because they enhance the folding rate of large protein molecules considerably, thus avoiding protein degradation of the nascent polypeptide chain (Jaenicke 1987).

There seems to be a structure-function relationship in connection with the organisation of the gene and gene families - the discovery of the exon-intron arrangement of enkaryotic genes suggested a correlation between the arrangement of such coding sequences in genes and protein domains, supporting ideas about their evolutionary origin. There is evidence that individual domains are coded by separate exons in multidomain proteins. Long, intronic DNA segments could serve the purpose of increasing the recombination rate among separate coding segments. In connection with protein divergence, this is a way in which different functional properties may have been reassorted to generate novel protein, thereby increasing the rate at which proteins evolved (= "exon shuffling")(Bajaj and Blundell 1984).

The functional importance of domains is evident: active sites of enzymes may reside on different domains or at domain interfaces; different substrates or effectors may bind to different domains performing different functions.

The assumption that the homeobox encoded a distinct protein domain is based upon the high degree of conservation observed in

many genes (as described earlier); where the homeobox does not occur as an exon in its own right, there is large variability in the flanking regions. Hence, it was postulated that the homeobox would give a homeodomain (Laughon and Scott 1984). The expression of a functional, structurally stable domain corresponding to the homeobox is borne out by the work in this thesis with the *ro*HD and that on the *Antp* HD (Muller et al.1988, Otting et al.1988, Qian et al.1989).

6.2 Modelling Homologous Proteins

The evolutionary mechanism of gene duplication and divergent evolution leads to families of homologous proteins that share similar primary and tertiary structures. This provides a powerful method of predicting the confirmation of one protein based on the crystal coordinates of a homologous protein. The process starts with the alignment of the two sequences during which insertions and deletions are generally placed in the loop regions connecting the regular secondary structural elements (as these tend to be more conserved during evolution). This is followed by model building (e.g. with interactive computer graphics) in which the known structure is initially taken, and alterations in the amino acid sequence are subsequently considered step-by-step. This starting model is then adjusted to yield a stereochemically sensible conformation in which i) there are no disallowed close contacts, ii) there is tight packing of the hydrophobic core, and iii) polar and charged atoms that do not form internal hydrogen bonds or salt bridges are exposed to the solvent.

Where the proteins have sequence homology of 50% or more, models have been predicted which are correct to better than 1A, although individual side chains may be more in error (Chothia 1984). Refinement of the model may be achieved by energy minimisation, e.g. using the CHARMM computer program (Shih et

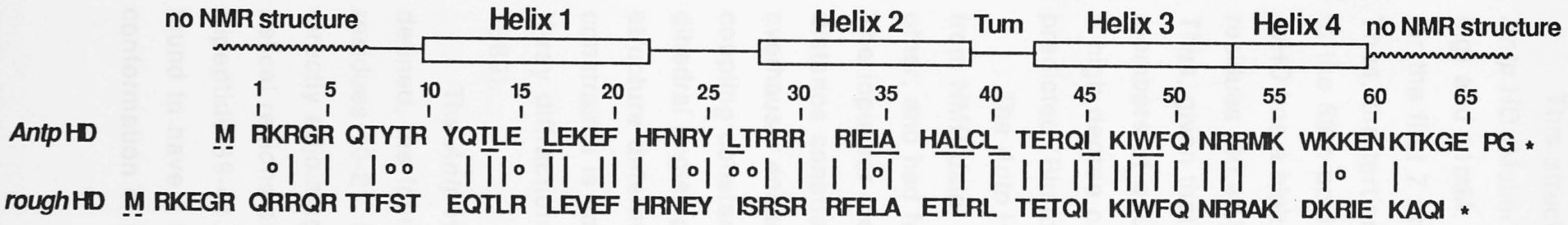
al.1985). In this procedure, the native conformation is found by moving the atoms in the model to give an equilibrium structure defined by a minimum of potential energy functions. However, there may by several different minimum energy conformations, and the native protein may not have the lowest potential energy conformation (since it may not be a kinetically accessible conformer on the protein folding pathway; Novotny et al.1984).

The tertiary structure of the *fushi tarazu* homeodomain has been modelled using sequence homology to the structurally-known DNA-binding proteins, catabolite activator protein (CAP) and λ repressor (Tsonis et al.1988) and subsequent energy minimisation to give the predicted structure. Although the helix-turn-helix segment was correctly predicted, there are significant structural differences when compared to the tertiary structure of the *Antp* HD (*ftz* and *Antp* are the same at 46/53 residues, and functionally conserved at 4 of the remaining 7, i.e. 94% homologous for the 53 residue NMR structure). For example, the two conserved interior residues Trp-48 and Phe-49 (Figures 6.2d) and e)) were predicted to be located on the protein surface (Qian et al.1989).

Although the tertiary structure of the *Antp* HD has been determined (Qian et al.1989), the three-dimensional coordinates corresponding to this structure had not been released at the time of writing. In the absence of this structural information a prediction of the *ro*HD structure could not be accurately modelled. However, a comparison of the primary sequence based on known functional similarities suggests that their structures are very closely related.

6.3 Tertiary Structure of the *Antennapedia* homeodomain

The secondary (Otting et al.1988) and tertiary structure (Qian et al.1989) of the *Antp* HD has been determined in solution by NMR spectroscopy.



The numbering of the *Antp* HD is from the first amino acid of the homeodomain, which finishes at residue 60; the M at the start of both the *Antp* HD and *rough*HD is an addition for expression. The *rough*HD has been aligned to show the degree of homology between the proteins: | = identical residues, o = residues which are functionally homologous. The three-dimensional structure of the *Antp* HD has been determined by NMR spectroscopy from residues 7 to 59 (Qian et al.1989/see Figure 6.2 for further details) and a summary of the secondary elements of the *Antp* HD is shown, including the helix 2-turn-helix 3 motif, whose existence had been proposed upon the basis of sequence homology with prokaryotic DNA-binding proteins (see Figure 4.2). The hydrophobic core of the *Antp* HD consists of the 11 underlined residues.

The *rough*HD has strong homology with the 53 residue portion of the *Antp* HD from 7 to 59 for which the NMR structure is known: 32/53 identical (60%), and 8/53 homologous (15%).

Figure 6.1: The sequence of the *Antp* HD protein (Muller et al.1988) compared with the *ro*HD protein (from pJG(2)); and details of the *Antp* HD protein's three-dimensional structure (Qian et al.1989).

This structural study was conducted with the 68 amino acid *Antp* HD (Muller et al.1988) at protein concentrations of 4 mM (13.5 mg) and 11mM. However, structural information was not obtained for the first 7 and last 8 residues of the protein (Qian et al.1989). Thus, the tertiary structure given is for 53 residues from the 7th to the 59th position in the *Antp* HD (Figure 6.1). In this region, the *ro*HD has a high homology with the *Antp* HD: 32/53 identical residues (60.4%), and 8/53 conserved residues (15.1%) (Figure 6.1). Thus, given that both the *ro*HD and *Antp* HD seem to represent members of a functionally and structurally conserved gene family, a high degree of structural similarity between the two would be predicted (Blundell et al.1987).

The *Antp* HD structure was based on 19 conformers derived from NMR data, all of these were in good agreement with each other, and had favourable conformation energies (Qian et al.1989). The input for these structure calculations contained about 7000 distance constraints determined by two-dimensional nuclear overhauser enhancement spectroscopy (NOESY) and 118 Scalar coupling constants that determine the allowed ranges for the dihedral angles phi and psi (Qian et al.1989). The quality of the structure determined from such an extensive set of structural constraints is considered to be comparable to that resulting from X-ray diffraction in single crystals at a resolution of ~2 A (Bax 1989).

The *Antp* HD contains four α-helices, three of these are well-defined, the fourth is more flexible (Figure 6.2). Helix 1 contains residues 10-21, helix 2 the residues 28-30, and helices 3 and 4 are directly adjoining from residues 42-52 and 53-59. These three helical regions are connected by the hexapeptide 22-27, and the tripeptide 39-41. Helix 1, 2, 3 are well-defined but helix 4 was found to have a less stable conformational structure. The conformation of the N- and C-terminal segments, 0-6 and 60-67,

Figure 6.2: Stereo views of the three-dimensional structure of the *Antp* HD (from Qian et al. 1989).

a) Protein backbone.

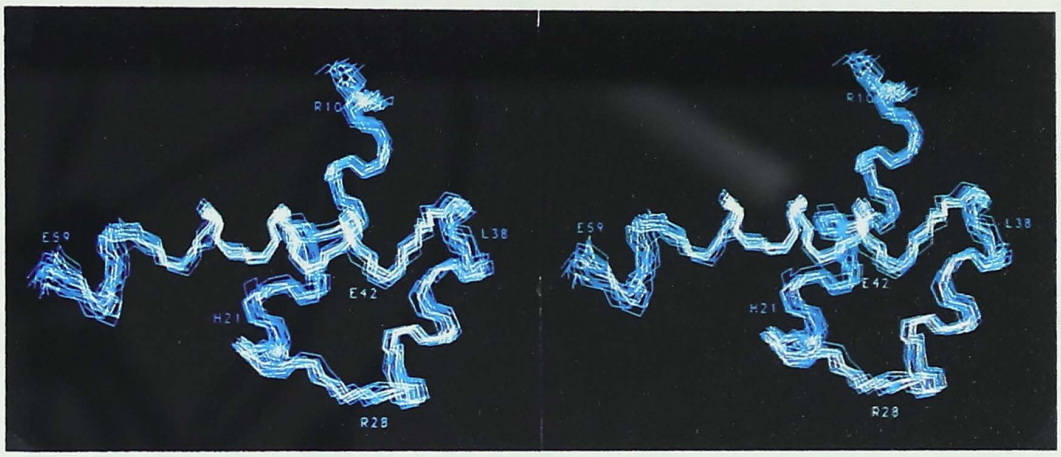
The polypeptide backbone atoms, N, C^{α}, and C' of residues 7 to 59 of the *Antp* HD (see Figure 6.1) are shown for the nineteen conformers resulting from distance geometry solutions of NMR data. Helix 1 is at the back, running down the middle from R10 to H21; the hexapeptide 22 to 27 runs along the bottom; helix 2 is on the right, running up from R28 to L38; after the tripeptide turn, helix 3 and helix 4 pass at the front, from right to left.

b) All heavy atoms.

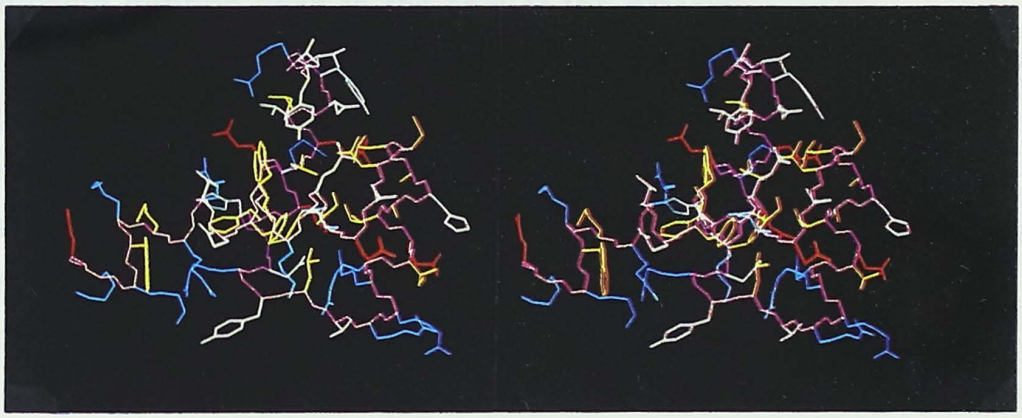
All of the heavy atoms (C,N,O,S) are shown for one of the conformers in a). The protein backbone is drawn in purple. Side chains are coloured: blue = Arg/Lys; red = Glu; yellow = Ala, Cys, Ile, Leu, Met, Phe, Trp; white = Asn, Gln, Ser, Thr, Tyr, His.

c) Hydrophobic core.

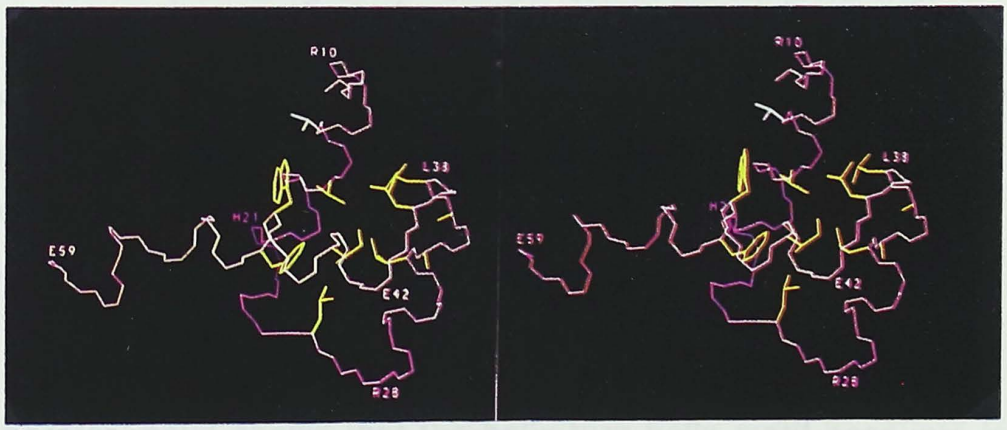
As in b), but only the backbone (purple), and the side chains comprising the hydrophobic core of the protein are shown, namely: Thr-13 (white), Leu-16, Leu-26, Ile-34, Ala-35, Ala-37, Leu-38, Leu-40, Ile-45, Trp-48, and Phe-49 (all in yellow).



a) Protein backbone.



b) All heavy atoms.



c) Hydrophobic core.

Figure 6.2: Stereo views of the three-dimensional structure of the *Antp* HD (from Qian et al. 1989).

was insufficiently defined by NMR to give a structure. However, this may be due to the presence in the *Antp* homeobox exon (the last of the four *Antp* exons; Schneuwly et al.1986) of an additional 15 amino acids at the C-terminus which may give a more stable conformation in the native protein, whilst at the N-terminus, where only one additional residue exists in the exon, a high degree of flexibility may be expected for the peptide link that connects the homeodomain with the rest of the native *Antp* protein. The *ro*HD is flanked by 9 (N-) and 8 (C-terminus) residues in the middle exon of *rough* and hence, would be predicted to have similar flexible peptide links to the rest of the protein.

The three-dimensional molecular architecture of the *Antp* HD (Figure 6.2a) shows that helices 1 and 2 are aligned in an almost exactly antiparallel fashion, whilst helices 3 and 4 are approximately perpendicular to 1 and 2. The axis of helix 4 forms an angle of about 30° with respect to the axis of helix 3 and it points away from the globular core of the protein.

Residues 30-50 form a helix-turn-helix motif in the *Antp* HD which is virtually identical to those observed in various prokaryotic repressors. This is the proposed DNA-binding region of the domain (see Chapter 4). Comparison of the *Antp* HD with several prokaryotic repressor proteins showed that, although the polypeptide backbone arrangement in the region of the helix-turn-helix segments is nearly identical in the *Antp* HD and phage 434 repressor, there are no additional similarities between the architectures of the two molecules. A high resolution crystal structure for the interaction of phage 434 repressor with a specific DNA region, and similar co-crystals of λ repressor-operator, have shown that the helix-turn-helix motif accounts for most of the sequence-specific DNA recognition and binding by these proteins (Aggarwal et al.1988, Jordan and Pabo 1988).

In all of the structures known for wild-type repressor

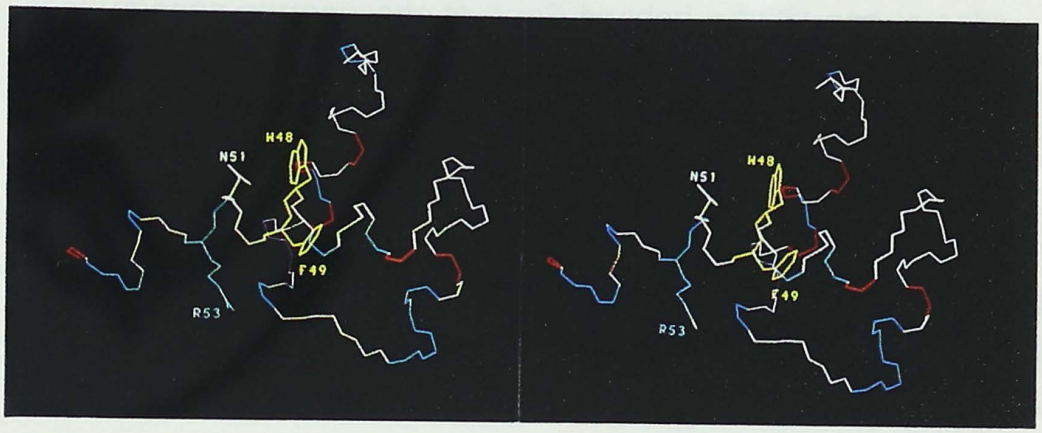
Figure 6.2: Stereo views of the three-dimensional structure of the *Antp* HD (from Qian et al. 1989).

d) Strictly conserved residues and charge distribution.

As in b) and c). Only the strictly conserved side chains are shown: Trp-48, Phe-49 (yellow); Asn-51 (white); and Arg-53 (blue). The backbone has been coloured blue for Arg/Lys, red for Glu, and purple for other residues.

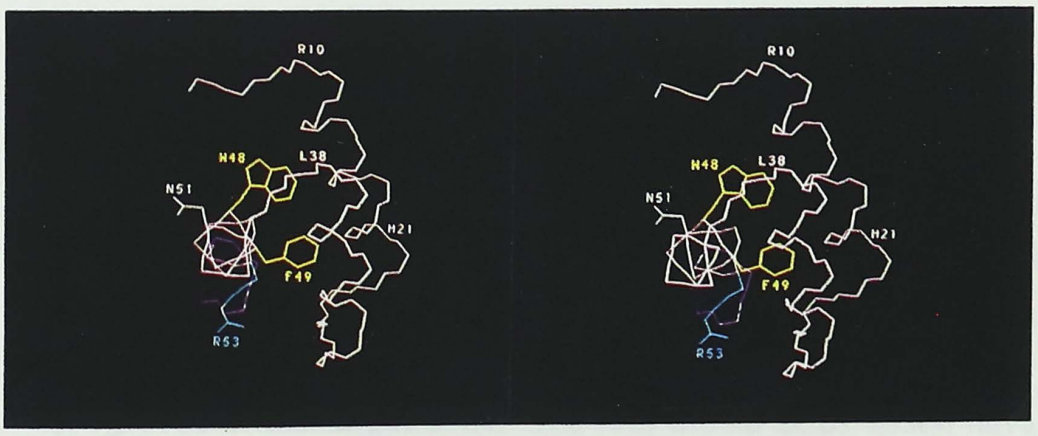
e) DNA-binding interactions.

The structure in b), c), and d) has been rotated to look down the axis of helix 3, the proposed recognition helix in DNA-binding interactions. The backbone is shown in purple. The strictly conserved side chains in d) are again shown:
Trp-48, Phe-49 (yellow); Asn-51 (white); and Arg-53 (blue).



d) Strictly conserved residues and charge distribution.

Figure 6.2: Stereo views of the three-dimensional structure of the *Antp* HD (from Qian et al. 1989).



e) DNA-binding interactions.

proteins, the first residue of the tripeptide forming the "turn" between the two helices is glycine, and it assumes a conformation defined by the dihedral angles, phi ~60°, psi ~60°. This conformation corresponds to a left-handed α-helix and is only energetically favourable for glycine. Only a few homeodomains (e.g. *engrailed,* Figure 4.2) have a glycine at this position, *Antp* has a cysteine and *rough*, an arginine; however, in the *Antp* HD, the cysteine at this position also has a local conformation with phi = 60° and psi = 60°. Some homeodomains have the sterically strongly restricted amino acids, valine and threonine, at this position. Such an energetically unfavourable conformation would presumably be countered by the force with which helices 1, 2 and 3 are held together. This is provided by the hydrophobic core, which in the *Antp* HD is formed by the 10 residues: Leu-16, Leu-26, Ile-34, Ala-35, Ala-37, Leu-38, Leu-40, Ile-45, Trp-48 and Phe-49. This protein core is further stabilised by a hydrogen bond between the side chain of Thr-13 and the indole nitrogen of Trp-48 (Figure 6.2c). In the *ro*HD, only one of these residues is notably different, with Thr at position 37 instead of Ala, hence it is likely that a similar hydrophobic core is formed.

The distribution of charged side chains in the *Antp* HD was noted, since it could contribute to favourable interactions of helix 3, and possibly helix 4, with DNA. Positive and negative charges are grouped on different sides of the protein (Figure 6.2d) - this asymmetric charge distribution should give the molecule a large global electric dipole moment. In addition, the *Antp* HD has a total of seven positive charges distributed along helices 3 and 4 (there are six in the *ro*HD, with Thr-43 instead of Arg) which could give energetically favourable interactions by these helices with the major groove of the target DNA. The four amino acids which have been found to be strictly conserved in the more than 80 homeodomains of metazoa (Scott et al.1989) are Trp-48 and Phe-

49, which are located in helix 3 of the *Antp* HD, where they are part of the hydrophobic core; and Asn-51 and Arg-53, which are located on the surface of helices 3 and 4 in *Antp* HD, and are thus likely candidates for interactions with DNA.

The 30° kink between helices 3 and 4 introduces a curvature into the helical segment from residues 42 to 59. There is speculation that this kink allows tight contacts to be made between the entire lengths of helices 3 and 4 and the target DNA. This could result in increased specificity and stability of the homeodomain-DNA complexes formed. Further information on the nature of the homeodomain's interactions with specific DNA sequences will presumably come from studies of co-crystals, although advances in 2-D NMR may permit a study in solution.

Bibliography.

Aggarwal A.K., Rodgers D.W., Drottar M., Ptashne M., Harrison S.C. (1988) Recognition of a DNA operator by the repressor of Phage 434 : a view at high resolution. Science 242 899-907.

Akam M. (1987) The molecular basis for metameric pattern in the *Drosophila* embryo. Development 101 1-22.

Anderson J.E. Ptashne M. Harrison S.C. (1987) Structure of the repressor-operator complex of Phage 434. Nature 326 846-52.

Armarego W.L.F. Cotton R.G.H. Dahl H.H.M. Dixon N.E. (1989) High level expression of human dihydropteridine reductase, without N-terminal protection, in *E. coli* Biochem. J. 261 265-68.

Bajaj M. and Blundell T. (1984) Evolution and the tertiary structure of proteins. Ann. Rev. Biophys. Bioeng. 13 453-92.

Banerjee U. Renfranz P.J. Pollock J.A. Benzer S. (1987) Molecular characterization and expression of *sevenless*, a gene involved in neuronal pattern formation in the *Drosophila* eye. Cell 49 281-91.

Bateson W. (1894) Materials for the Study of Variation, Macmillan & Co. London.

Bax A. (1989) Two-dimensional NMR and protein structure. Ann. Rev. Biochem. 58 223-56.

Beachy P.A. Krasnow M.A. Gavis E.R. Hogness D.S. (1988) An *Ultrabithorax* protein binds sequences near its own and the *Antennapedia* P1 promoters. Cell 55 1069-81.

Berger S.L. and Kimmel A.R. (1987) Guide to molecular cloning techniques. Meths. in Enzym. 152.

Blundell T.L. Sibanda B.L. Sternberg M.J.E. Thornton J.M. (1987)

Knowledge-based prediction of protein structures and the design of novel molecules. Nature 326 347-52.

Bodner M. Castrillo J.L. Thiell L.E. Deerinck T. Ellisman M. Karin M. (1988) The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell 55 505-18.

Bowen B. Steinberg J. Laemmli U.K. Weintraub H. (1980) The detection of DNA-binding proteins by protein blotting. Nuc. Ac. Res. 8 1-20.

Buell G. and Panayotatos N., 1986. Mechanism and Practice, p.345-364 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Caroll S.B. and Scott M.P. (1985) Localization of the *fushi tarazu* protein during *Drosophila* embryogenesis. Cell 43 47-57.

Chambers S.P. Prior S.E. Barstow D.A. Minton N.P. (1988) The pMTL nic$^-$ cloning vectors . i. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleptide sequencing. Gene 68 139-49.

Cho K.W.Y. Goetz J. Wright C.V.E. Fritz A. Hardwicke J. De Robertis E.M. (1988) Differential utilization of the same reading frame in a *Xenopus* homeobox gene encodes two related proteins sharing the same DNA-binding specificity. EMBO J. 7 2139-49.

Chothia C. (1984) Principles that determine the structure of proteins. Ann. Rev. Biochem. 53 537-72.

Dearolf C.R. Topol J. Parker C.S. (1989) The *caudal* gene product is a direct activator of *fushi tarazu* transcription during *Drosophila* embryogenesis. Nature 341 340-43.

Desplan C. Theis J. O'Farrell P.H. (1985) The *Drosophila* developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity Nature 318 630-635.

Desplan C. Theis J. O'Farrell P.H. (1988) The sequence-specificity

of homeodomain-DNA interaction. Cell 54 1081-90.

DeWald D.B. Adams L.D. Pearson J.D. (1986) A non-urea gel system for resolution of M_r 2000 to 200,000. Anal. Bioch. 154 502-8.

Dillon J.A. Nasim A. Nestmann E.R. (1985) Recombinant DNA methodology Wiley N.Y.

DiNardo S. Kuner J.M. Theis J. O'Farrell P.H. (1985) Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. Cell 43 59-69.

Drew H.R. and Travers A.A. (1985) Structural junctions in DNA : the influence of flanking sequence on nuclease digestion specificities. Nuc. Ac. Res. 13 4445-67.

Drew H.R. and McCall M.J. (1988) Recent studies of DNA in the crystal. Ann. Rev. Cell Biol. 4 1-20.

Driever W. Ma J. Nusslein-Volhard C. Ptashne M. (1989) Rescue of *bicoid* mutant *Drosophila* embryos by Bicoid fusion proteins containing heterologous activating sequences. Nature 342 149-54.

Elvin C.M. Thompson P.R. Argall M.E. Hendry P. Stamford N.P.J. Lilley P.E. Dixon N.E. (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *E. coli*. Gene (in press).

Fried M. and Crothers D.M. (1981) Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nuc. Ac. Res. 23 6505-25.

Galas D.J. and Schmitz A. (1978) DNAase footprinting : a simple method for the detection of protein-DNA binding specificity. Nuc. Ac. Res. 5 3157-70.

Garner M.M. and Revzin A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions : application to components of the *E. coli* lactose

operon regulatory system. Nuc. Ac. Res. 9 3047-60.

Garner M.M. and Revzin A. (1986) The use of gel electrophoresis to detect and study nucleic acid-protein interactions. TIBS 11 395-6.

Guilian G.G. Moss R.L. Greaser M. (1984) Analytical isoelectric focusing using a high-voltage vertical slab polacrylamide gel system. Anal. Bioch. 142 421-436.

Go N. (1983) Theoretical studies of protein folding. Ann. Rev. Biophys. Bioeng. 12 183-210.

Goff S.A. Casson L. Goldberg A. (1984) Heat Shock Regulatory gene *htpR* influences rates of protein degradation and expression of the *lon* gene in *E. Coli.* P.N.A.S. 81 6647-51.

Goldberg A.L. and Goff S.A. (1986) The Selective Degradation of Abnormal Proteins in Bacteria. p.287-314 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Golonbinoff P. Christeller J.T. Gatenby A.A. Lorimer G.H. (1989) Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. Nature 342 884-89.

Gottesman M.E. Adhya S. Das A. (1980) Transcriptional antitermination by bacteriophage λ *N* gene product. J. Mol. Biol. 140 57-75.

Han K. Levine M.S. Manley J.L. (1989) Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. Cell 56 573-83.

Hanes S.D. and Brent R. (1989) DNA Specificity of the Bicoid Activator Protein is Determined by homeodomain Recognition Helix Residue 9. Cell 57 1275-1283.

He X. Treacy M.N. Simmons D.M. Ingraham H.A. Swanson L.W. Rosenfeld M.G. (1989) Expression of a large family of POU-domain regulatory genes in mammalian brain development.

Nature 340 35-42.

Hendrickson W. (1985) Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay. BioTechniques 5 198-207.

Hennighausen L. and Lubon H. (1987) Interaction of protein with DNA *in vitro*. Meth. in Enzym. 152 721-35.

Hoey T. and Levine M. (1988) Divergent homeobox proteins recognise similar DNA sequences in *Drosophila*. Nature 332 858-61.

Howell M. J. and Hargreaves J.J. (1988) Cloning and expression of *Taenia oris* antigens in *E. Coli.* Mol. Bioch. Parasit. 28 21-30.

Hubscher U. (1987) Double replica Southwestern. Nuc. Ac. Res. 15 5486.

Jaenicke R. (1987) Folding and Association of Proteins. Progress in Biophysics and Molecular Biology 49 117-237.

Jordan S.R. and Pabo C.O. (1988) Structure of the lambda complex at 2.5 A resolution : details of the repressor-operator interactions. Science 242 893-98.

Kaufmann T.C. Lewis R. Wakimoto, B. (1980) Cytogenetic analysis of chromosome 3 in *D. melanogaster* : the homeotic gene complex in polytene chromosome interval 84A,B. Genetics 94 115-133.

Kennell D.E. (1986) The Instability of Messenger RNA in Bacteria. p.101-142 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Kuziora M.A. and McGinnis W. (1989) A homeodomain substitution changes the regulatory specificity of the *Deformed* protein in *Drosophila* embryos. Cell 59 563-71.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 680-85.

Laughon A. and Scott M.P. (1984) Sequence of a *Drosophila* Segmentation gene: protein structure homology with DNA-binding proteins. Nature 310 25-31.

Lewis E.B. (1978) A gene complex controlling segmentation in *Drosophila* . Nature 276 565-70.

Maniatis T. Fritsch E.F. Sambrook J. (1982) Molecular cloning : a laboratory manual. CSHLab.CSHN.Y.

Maniatis T. Goodbourn S. Fischer J.A. (1987) Regulation of inducible and tissue-specific gene expression. Science 236 1237-44.

Matsudiara P. (1987) Sequence from picomole quantities of proteins electroblotted onto PVDF membranes. J. Biol. Chem. 10035-38.

Maxam A.M. and Gilbert W. (1977) A new method for sequencing DNA. PNAS 74 560-64.

McGinnis W. Levine M. Hafen E. Kuroiwa A. Gehring W.J. (1984a) A Conserved DNA sequence found in Homeotic Genes of the *Drosophila Antennapedia*. and *Bithorax* Complexes. Nature 308 428-33.

McGinnis W. Garber R.L. Wirz J. Kuroiwa A. Gehring W.J. (1984b) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell 37 403-408.

Mihara H. and Kaiser E.T. (1988) Achemically synthesized *Antennapedia* homeodomain binds to a specific DNA sequence. Science 242 925-27.

Muller M. Affolter M. Leupin W. Otting G. Wuthrich K. Gehring W.J. (1988) Isolation and sequence-specific DNA binding of the *Antennapedia* homeodomain. EMBO J. 7 4299-4304.

Muller M.M. Ruppert S. Schaffner W. Matthias P. (1988) A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature 336 544-557.

Novotny J. Bruccoleri R. Karplus M. (1984) An analysis of incorrectly folded protein models. J. Mol. Biol. 177 787-818.

O'Shea E.K. Rutkowski R. Kim P.S. (1989) Evidence that the Leucine Zipper is a coiled coil. Science 243 538-42.

Otting G. Qian Y. Muller M. Affolter M. Gehring W.J. Wuthrich K. (1988) Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. EMBO J. 7 4305-4309.

Pabo C.O. and Sauer R.T. (1984) Protein-DNA Recognition. Ann. Rev. Biochem. 53 293-321.

Panyim S. and Chalkley R. (1971) The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method. J. Biol. Chem. 246 7557-60.

Ptashne M. Johnson A.D. Pabo C.O. (1982) A Genetic Switch in a Bacterial Virus. Sci. Am. 247 128-40.

Ptashne M. (1986) Gene regulation by proteins acting nearby and at a distance. Nature 322 697-701.

Ptashne M. (1988) How eukaryotic transcriptional activators work. Nature 335 683-89.

Qian Y.A. Billeter M. Otting G. Muller M. Gehring W.J. Wuthrich K. (1989) The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution : comparison with prokaryotic repressors. Cell 59 573-80.

Riggs A.D. Suzuki H. Bourgeois S. (1970) *lac* repressor-operator interaction. I. Equilibrium studies. J. Mol. Biol. 48 67-83.

Rosenfeld P.J. and Kelly T.J. (1986) Purification of Nuclear Factor I by DNA recognition site affinity chromatography. J. Biol. Chem. 261 1398-1408.

Rouviere-Yaniv J. and Gros F. (1975) Characterisation of a novel, low-molecular-weight DNA-binding protein from *E. coli*.

PNAS 72 3428-3432

Rubin G.M. (1989) Development of the *Drosophila* retina : inductive events studied at single cell resolution. Cell 57 519-20.

Saint R. Kalionis B. Lockett T.J. Elizur A. (1988) Pattern formation in the developing eye of *Drosophila melanogaster* is regulated by the homeobox gene, *rough.* Nature 334 151-54.

Sanger F. Coulson A.R. Hong G.F. Hill D.F. Petersen G.B. (1982) Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. 162 729-73.

Scheidereit C. Cromlish J.A. Gerster T. Kawakami K. Balmaceda C. Currie R.A. Roeder R.G. (1988) A Human Lymphoid-Specific Transcription Factor that Activates Immunoglobulin Genes is a Homoeobox Protein. Nature 336 551-557.

Schlief R. (1988) DNA Binding by Proteins. Science 241 1182-87.

Schneuwly S. Kuriowa A. Baumgartner P. Gehring W.J. (1986) Structural Organisation and sequence of the Homeotic Gene *Antennapedia* of *Drosophila.* EMBO J. 5 733-39.

Scott M.P. Tamkun J.W. Hartzell G.W. (1989) The Structure and Function of the Homeodomain. Bioch. et Biophys. Acta 989 25-48.

Seeman W.C. Rosenberg J.M. Rich A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. PNAS 73 804-8.

Shih H.H.L. Brady J. Karplus M. (1985) Structure of proteins with single-site mutations : a minimum perturbation approach. PNAS 82 1697-1700.

Shine J. and Dalgarno L. (1974) The 3'-terminal end of *E. coli* 16S ribosomal RNA : complementarity to nonsense triplets and ribosome binding sites. PNAS 71 1342-1346.

Silva C.M. Tully D.B. Petch L.A. Jewell C.M. Cidlowski J.A. (1987) Application of a protein-blotting procedure to the study of human glucocorticoid receptor interactions with DNA. PNAS

84 1744-48.

Sorger P.K. Ammarer G. Shore D. (1989) Identification and purification of sequence-specific DNA-binding proteins. p.199-223 in Protein Function Creighton T. ed. IRL Press Oxford.

Stanley K.K. (1983) Solubilisation and immune-detection of β-galactosidase hybrid proteins carrying foreign antigenic determinants. Nuc. Ac. Res. 11 4077-92.

Stanley K.K. and Luzio J.P. (1984) Construction of a new family of high efficiency bacterial expression vectors : identification of cDNA clones coding for human liver proteins. EMBO J. 3 1429-34.

Stormo G.D. (1986) Translation Initiation. p.195-224 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Sturm R. Das G. Herr W. (1988) The POU Domain is a bipartite DNA-binding structure. Nature 336 601-4.

Szewczyk B. and Kozloff L.M. (1985) A method for the efficient blotting of strongly basic proteins from sodium dodecyl sulfate-polyacrylamide gels to nitrocellulose. Anal. Bioch. 150 403-7.

Thali M. Müller M.M. DeLorenzi M. Matthias P. Bienz M. (1988) *Drosophila* Homoeotic Genes encode Transcriptional Activators similar to Mammalian OTF-2 Nature 336 598-601.

Tomlinson A. and Ready D.F. (1987) Neuronal differentiation in the *Drosophila* ommatidium. Dev. Biol. 120 366-76.

Tomlinson A. Kimmel B. Rubin G. (1988a) *Rough*, a *Drosophila* Homeobox Gene Required in Photoreceptors R2 and R5 for Inductive Interactions in the Developing Eye. Cell 55 771-784.

Tomlinson A. (1988b) Cellular interactions in the developing

Drosophila eye. Development 104 183-93.

Tora L. White J. Brou C. Tasset D. Webster N. Scheer E. Chambon P. (1989) The human Estrogen Receptor has two independent non-acidic transcriptional activation functions. Cell 59 477-87.

Towbin H. Staehelin T. Gordon J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets : procedure and some applications. PNAS 76 4350-54.

Travers A.A. and Klug A.(1987) The bending of DNA in nucleosomes and its wider implications. Phil. Trans. R. Soc. Lond. B 317 537-61.

Travers A.A. (1989) DNA Conformation and Protein Binding. Ann. Rev. Biochem. 58 427-52.

Treisman J. and Desplan C. (1989) The products of the *Drosophila* gap genes Hunchback and Kruppel bind to the *hunchback* promoters. Nature 341 355-58.

Treisman J. Gonczy P. Vashishtha M. Harris E. Desplan C. (1989) A single amino acid can determine the DNA binding specificity of homeodomaion proteins. Cell 59 553-62.

Tsonis P.A. Carperos V. Shiahaan T. (1988) Modelling of the homeodomain suggests similar structure to repressors. Bioch. Biophys. Res. Comm. 157 100-5.

Tullius T.D. (1989) Physical studies of protein-DNA complexes by footprinting. Ann. Rev. Biophys. Biophys. Chem. 18 213-37.

Vieira J. and Messing J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagensis and sequencing with synthetic universal primers. Gene 19 259-68.

Weber K. and Osborn M. (1975) Proteins and sodium dodecyl sulfate : molecular weight determination on polyacrylamide gels and related procedures. p.179-191 in The Proteins Vol. I Neurath H./ Hill R.L.(eds.) Acad. Press N.Y.

Wharton K.A. Yedvobnick B. Finnerty V.G. Artavanis-Tsakonas S. (1985) Opa: A Novel Family of Transcribed Repeats Shared by the Notch Locus and other Developmentally Regulated Loci in *D. Melanogaster*. Cell 40 55-62.

Wharton R.P. and Ptashne M. (1985) Changing the binding specificity of a repressor by redesigning an α-helix. Nature 316 601-5.

White R.A.H. and Wilcox M. (1984) Protein products of the Bithorax complex in *Drosophila*. Cell 39 163-71.

Williams D.C. Van Frank R.M. Muth W.L. Burnett P.J. (1982) Cytoplasmic Inclusion Bodies in *E. Coli* Producing Biosynthetic Human Insulin Proteins. Science 215 687-689.

Yanisch-Peron C. Vieira J. Messing J. (1985) Improved M13 phage vectors and host strains : nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 103-19.

Bibliography.

Aggarwal A.K., Rodgers D.W., Drottar M., Ptashne M., Harrison S.C. (1988) Recognition of a DNA operator by the repressor of Phage 434 : a view at high resolution. Science 242 899-907.

Akam M. (1987) The molecular basis for metameric pattern in the *Drosophila* embryo. Development 101 1-22.

Anderson J.E. Ptashne M. Harrison S.C. (1987) Structure of the repressor-operator complex of Phage 434. Nature 326 846-52.

Armarego W.L.F. Cotton R.G.H. Dahl H.H.M. Dixon N.E. (1989) High level expression of human dihydropteridine reductase, without N-terminal protection, in *E. coli* Biochem. J. 261 265-68.

Bajaj M. and Blundell T. (1984) Evolution and the tertiary structure of proteins. Ann. Rev. Biophys. Bioeng. 13 453-92.

Banerjee U. Renfranz P.J. Pollock J.A. Benzer S. (1987) Molecular characterization and expression of *sevenless*, a gene involved in neuronal pattern formation in the *Drosophila* eye. Cell 49 281-91.

Bateson W. (1894) Materials for the Study of Variation, Macmillan & Co. London.

Bax A. (1989) Two-dimensional NMR and protein structure. Ann. Rev. Biochem. 58 223-56.

Beachy P.A. Krasnow M.A. Gavis E.R. Hogness D.S. (1988) An *Ultrabithorax* protein binds sequences near its own and the *Antennapedia* P1 promoters. Cell 55 1069-81.

Berger S.L. and Kimmel A.R. (1987) Guide to molecular cloning techniques. Meths. in Enzym. 152.

Blundell T.L. Sibanda B.L. Sternberg M.J.E. Thornton J.M. (1987)

Knowledge-based prediction of protein structures and the design of novel molecules. Nature 326 347-52.

Bodner M. Castrillo J.L. Thiell L.E. Deerinck T. Ellisman M. Karin M. (1988) The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell 55 505-18.

Bowen B. Steinberg J. Laemmli U.K. Weintraub H. (1980) The detection of DNA-binding proteins by protein blotting. Nuc. Ac. Res. 8 1-20.

Buell G. and Panayotatos N., 1986. Mechanism and Practice, p.345-364 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Caroll S.B. and Scott M.P. (1985) Localization of the *fushi tarazu* protein during *Drosophila* embryogenesis. Cell 43 47-57.

Chambers S.P. Prior S.E. Barstow D.A. Minton N.P. (1988) The pMTL nic$^-$ cloning vectors . i. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleptide sequencing. Gene 68 139-49.

Cho K.W.Y. Goetz J. Wright C.V.E. Fritz A. Hardwicke J. De Robertis E.M. (1988) Differential utilization of the same reading frame in a *Xenopus* homeobox gene encodes two related proteins sharing the same DNA-binding specificity. EMBO J. 7 2139-49.

Chothia C. (1984) Principles that determine the structure of proteins. Ann. Rev. Biochem. 53 537-72.

Dearolf C.R. Topol J. Parker C.S. (1989) The *caudal* gene product is a direct activator of *fushi tarazu* transcription during *Drosophila* embryogenesis. Nature 341 340-43.

Desplan C. Theis J. O'Farrell P.H. (1985) The *Drosophila* developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity Nature 318 630-635.

Desplan C. Theis J. O'Farrell P.H. (1988) The sequence-specificity

of homeodomain-DNA interaction. Cell 54 1081-90.

DeWald D.B. Adams L.D. Pearson J.D. (1986) A non-urea gel system for resolution of M_r 2000 to 200,000. Anal. Bioch. 154 502-8.

Dillon J.A. Nasim A. Nestmann E.R. (1985) Recombinant DNA methodology Wiley N.Y.

DiNardo S. Kuner J.M. Theis J. O'Farrell P.H. (1985) Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. Cell 43 59-69.

Drew H.R. and Travers A.A. (1985) Structural junctions in DNA : the influence of flanking sequence on nuclease digestion specificities. Nuc. Ac. Res. 13 4445-67.

Drew H.R. and McCall M.J. (1988) Recent studies of DNA in the crystal. Ann. Rev. Cell Biol. 4 1-20.

Driever W. Ma J. Nusslein-Volhard C. Ptashne M. (1989) Rescue of *bicoid* mutant *Drosophila* embryos by Bicoid fusion proteins containing heterologous activating sequences. Nature 342 149-54.

Elvin C.M. Thompson P.R. Argall M.E. Hendry P. Stamford N.P.J. Lilley P.E. Dixon N.E. (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *E. coli*. Gene (in press).

Fried M. and Crothers D.M. (1981) Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nuc. Ac. Res. 23 6505-25.

Galas D.J. and Schmitz A. (1978) DNAase footprinting : a simple method for the detection of protein-DNA binding specificity. Nuc. Ac. Res. 5 3157-70.

Garner M.M. and Revzin A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions : application to components of the *E. coli* lactose

operon regulatory system. Nuc. Ac. Res. 9 3047-60.

Garner M.M. and Revzin A. (1986) The use of gel electrophoresis to detect and study nucleic acid-protein interactions. TIBS 11 395-6.

Guilian G.G. Moss R.L. Greaser M. (1984) Analytical isoelectric focusing using a high-voltage vertical slab polacrylamide gel system. Anal. Bioch. 142 421-436.

Go N. (1983) Theoretical studies of protein folding. Ann. Rev. Biophys. Bioeng. 12 183-210.

Goff S.A. Casson L. Goldberg A. (1984) Heat Shock Regulatory gene *htpR* influences rates of protein degradation and expression of the *lon* gene in *E. Coli.* P.N.A.S. 81 6647-51.

Goldberg A.L. and Goff S.A. (1986) The Selective Degradation of Abnormal Proteins in Bacteria. p.287-314 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Golonbinoff P. Christeller J.T. Gatenby A.A. Lorimer G.H. (1989) Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. Nature 342 884-89.

Gottesman M.E. Adhya S. Das A. (1980) Transcriptional antitermination by bacteriophage λ *N* gene product. J. Mol. Biol. 140 57-75.

Han K. Levine M.S. Manley J.L. (1989) Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. Cell 56 573-83.

Hanes S.D. and Brent R. (1989) DNA Specificity of the Bicoid Activator Protein is Determined by homeodomain Recognition Helix Residue 9. Cell 57 1275-1283.

He X. Treacy M.N. Simmons D.M. Ingraham H.A. Swanson L.W. Rosenfeld M.G. (1989) Expression of a large family of POU-domain regulatory genes in mammalian brain development.

Nature 340 35-42.

Hendrickson W. (1985) Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay. BioTechniques 5 198-207.

Hennighausen L. and Lubon H. (1987) Interaction of protein with DNA *in vitro*. Meth. in Enzym. 152 721-35.

Hoey T. and Levine M. (1988) Divergent homeobox proteins recognise similar DNA sequences in *Drosophila*. Nature 332 858-61.

Howell M. J. and Hargreaves J.J. (1988) Cloning and expression of *Taenia oris* antigens in *E. Coli.* Mol. Bioch. Parasit. 28 21-30.

Hubscher U. (1987) Double replica Southwestern. Nuc. Ac. Res. 15 5486.

Jaenicke R. (1987) Folding and Association of Proteins. Progress in Biophysics and Molecular Biology 49 117-237.

Jordan S.R. and Pabo C.O. (1988) Structure of the lambda complex at 2.5 A resolution : details of the repressor-operator interactions. Science 242 893-98.

Kaufmann T.C. Lewis R. Wakimoto, B. (1980) Cytogenetic analysis of chromosome 3 in *D. melanogaster* : the homeotic gene complex in polytene chromosome interval 84A,B. Genetics 94 115-133.

Kennell D.E. (1986) The Instability of Messenger RNA in Bacteria. p.101-142 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Kuziora M.A. and McGinnis W. (1989) A homeodomain substitution changes the regulatory specificity of the *Deformed* protein in *Drosophila* embryos. Cell 59 563-71.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 680-85.

Laughon A. and Scott M.P. (1984) Sequence of a *Drosophila* Segmentation gene: protein structure homology with DNA-binding proteins. Nature 310 25-31.

Lewis E.B. (1978) A gene complex controlling segmentation in *Drosophila* . Nature 276 565-70.

Maniatis T. Fritsch E.F. Sambrook J. (1982) Molecular cloning : a laboratory manual. CSHLab.CSHN.Y.

Maniatis T. Goodbourn S. Fischer J.A. (1987) Regulation of inducible and tissue-specific gene expression. Science 236 1237-44.

Matsudiara P. (1987) Sequence from picomole quantities of proteins electroblotted onto PVDF membranes. J. Biol. Chem. 10035-38.

Maxam A.M. and Gilbert W. (1977) A new method for sequencing DNA. PNAS 74 560-64.

McGinnis W. Levine M. Hafen E. Kuroiwa A. Gehring W.J. (1984a) A Conserved DNA sequence found in Homeotic Genes of the *Drosophila Antennapedia*. and *Bithorax* Complexes. Nature 308 428-33.

McGinnis W. Garber R.L. Wirz J. Kuroiwa A. Gehring W.J. (1984b) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell 37 403-408.

Mihara H. and Kaiser E.T. (1988) Achemically synthesized *Antennapedia* homeodomain binds to a specific DNA sequence. Science 242 925-27.

Muller M. Affolter M. Leupin W. Otting G. Wuthrich K. Gehring W.J. (1988) Isolation and sequence-specific DNA binding of the *Antennapedia* homeodomain. EMBO J. 7 4299-4304.

Muller M.M. Ruppert S. Schaffner W. Matthias P. (1988) A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature 336 544-557.

Novotny J. Bruccoleri R. Karplus M. (1984) An analysis of incorrectly folded protein models. J. Mol. Biol. 177 787-818.

O'Shea E.K. Rutkowski R. Kim P.S. (1989) Evidence that the Leucine Zipper is a coiled coil. Science 243 538-42.

Otting G. Qian Y. Muller M. Affolter M. Gehring W.J. Wuthrich K. (1988) Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. EMBO J. 7 4305-4309.

Pabo C.O. and Sauer R.T. (1984) Protein-DNA Recognition. Ann. Rev. Biochem. 53 293-321.

Panyim S. and Chalkley R. (1971) The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method. J. Biol. Chem. 246 7557-60.

Ptashne M. Johnson A.D. Pabo C.O. (1982) A Genetic Switch in a Bacterial Virus. Sci. Am. 247 128-40.

Ptashne M. (1986) Gene regulation by proteins acting nearby and at a distance. Nature 322 697-701.

Ptashne M. (1988) How eukaryotic transcriptional activators work. Nature 335 683-89.

Qian Y.A. Billeter M. Otting G. Muller M. Gehring W.J. Wuthrich K. (1989) The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution : comparison with prokaryotic repressors. Cell 59 573-80.

Riggs A.D. Suzuki H. Bourgeois S. (1970) *lac* repressor-operator interaction. I. Equilibrium studies. J. Mol. Biol. 48 67-83.

Rosenfeld P.J. and Kelly T.J. (1986) Purification of Nuclear Factor I by DNA recognition site affinity chromatography. J. Biol. Chem. 261 1398-1408.

Rouviere-Yaniv J. and Gros F. (1975) Characterisation of a novel, low-molecular-weight DNA-binding protein from *E. coli*.

PNAS 72 3428-3432

Rubin G.M. (1989) Development of the *Drosophila* retina : inductive events studied at single cell resolution. Cell 57 519-20.

Saint R. Kalionis B. Lockett T.J. Elizur A. (1988) Pattern formation in the developing eye of *Drosophila melanogaster* is regulated by the homeobox gene, *rough.* Nature 334 151-54.

Sanger F. Coulson A.R. Hong G.F. Hill D.F. Petersen G.B. (1982) Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. 162 729-73.

Scheidereit C. Cromlish J.A. Gerster T. Kawakami K. Balmaceda C. Currie R.A. Roeder R.G. (1988) A Human Lymphoid-Specific Transcription Factor that Activates Immunoglobulin Genes is a Homoeobox Protein. Nature 336 551-557.

Schlief R. (1988) DNA Binding by Proteins. Science 241 1182-87.

Schneuwly S. Kuriowa A. Baumgartner P. Gehring W.J. (1986) Structural Organisation and sequence of the Homeotic Gene *Antennapedia* of *Drosophila.* EMBOJ.5 733-39.

Scott M.P. Tamkun J.W. Hartzell G.W. (1989) The Structure and Function of the Homeodomain. Bioch. et Biophys. Acta 989 25-48.

Seeman W.C. Rosenberg J.M. Rich A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. PNAS 73 804-8.

Shih H.H.L. Brady J. Karplus M. (1985) Structure of proteins with single-site mutations : a minimum perturbation approach. PNAS 82 1697-1700.

Shine J. and Dalgarno L. (1974) The 3'-terminal end of *E. coli* 16S ribosomal RNA : complementarity to nonsense triplets and ribosome binding sites. PNAS 71 1342-1346.

Silva C.M. Tully D.B. Petch L.A. Jewell C.M. Cidlowski J.A. (1987) Application of a protein-blotting procedure to the study of human glucocorticoid receptor interactions with DNA. PNAS

84 1744-48.

Sorger P.K. Ammarer G. Shore D. (1989) Identification and purification of sequence-specific DNA-binding proteins. p.199-223 in Protein Function Creighton T. ed. IRL Press Oxford.

Stanley K.K. (1983) Solubilisation and immune-detection of β-galactosidase hybrid proteins carrying foreign antigenic determinants. Nuc. Ac. Res. 11 4077-92.

Stanley K.K. and Luzio J.P. (1984) Construction of a new family of high efficiency bacterial expression vectors : identification of cDNA clones coding for human liver proteins. EMBO J. 3 1429-34.

Stormo G.D. (1986) Translation Initiation. p.195-224 in Maximising gene expression Reznikoff, W./Gold, L. (eds.) Butterworths, Massach. 1986.

Sturm R. Das G. Herr W. (1988) The POU Domain is a bipartite DNA-binding structure. Nature 336 601-4.

Sussman R. and Jacob F. (1962) Sur un systeme de repression thermosensible chez le bacteriophage d'*E. coli.* C. R. Acad. Sci. 254 1517.

Szewczyk B. and Kozloff L.M. (1985) A method for the efficient blotting of strongly basic proteins from sodium dodecyl sulfate-polyacrylamide gels to nitrocellulose. Anal. Bioch. 150 403-7.

Thali M. Müller M.M. DeLorenzi M. Matthias P. Bienz M. (1988) *Drosophila* Homoeotic Genes encode Transcriptional Activators similar to Mammalian OTF-2 Nature 336 598-601.

Tomlinson A. and Ready D.F. (1987) Neuronal differentiation in the *Drosophila* ommatidium. Dev. Biol. 120 366-76.

Tomlinson A. Kimmel B. Rubin G. (1988a) *Rough*, a *Drosophila* Homeobox Gene Required in Photoreceptors R2 and R5 for Inductive Interactions in the Developing Eye. Cell 55 771-784.

Tomlinson A. (1988b) Cellular interactions in the developing *Drosophila* eye. Development 104 183-93.

Tora L. White J. Brou C. Tasset D. Webster N. Scheer E. Chambon P. (1989) The human Estrogen Receptor has two independent non-acidic transcriptional activation functions. Cell 59 477-87.

Towbin H. Staehelin T. Gordon J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets : procedure and some applications. PNAS 76 4350-54.

Travers A.A. and Klug A.(1987) The bending of DNA in nucleosomes and its wider implications. Phil. Trans. R. Soc. Lond. B 317 537-61.

Travers A.A. (1989) DNA Conformation and Protein Binding. Ann. Rev. Biochem. 58 427-52.

Treisman J. and Desplan C. (1989) The products of the *Drosophila* gap genes Hunchback and Kruppel bind to the *hunchback* promoters. Nature 341 355-58.

Treisman J. Gonczy P. Vashishtha M. Harris E. Desplan C. (1989) A single amino acid can determine the DNA binding specificity of homeodomaion proteins. Cell 59 553-62.

Tsonis P.A. Carperos V. Shiahaan T. (1988) Modelling of the homeodomain suggests similar structure to repressors. Bioch. Biophys. Res. Comm. 157 100-5.

Tullius T.D. (1989) Physical studies of protein-DNA complexes by footprinting. Ann. Rev. Biophys. Biophys. Chem. 18 213-37.

Vieira J. and Messing J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagensis and sequencing with synthetic universal primers. Gene 19 259-68.

Weber K. and Osborn M. (1975) Proteins and sodium dodecyl sulfate : molecular weight determination on polyacrylamide gels and related procedures. p.179-191 in The Proteins Vol. I Neurath H./ Hill R.L.(eds.) Acad. Press N.Y.

Wharton K.A. Yedvobnick B. Finnerty V.G. Artavanis-Tsakonas S. (1985) Opa: A Novel Family of Transcribed Repeats Shared by the Notch Locus and other Developmentally Regulated Loci in *D. Melanogaster*. Cell 40 55-62.

Wharton R.P. and Ptashne M. (1985) Changing the binding specificity of a repressor by redesigning an α-helix. Nature 316 601-5.

White R.A.H. and Wilcox M. (1984) Protein products of the Bithorax complex in *Drosophila*. Cell 39 163-71.

Williams D.C. Van Frank R.M. Muth W.L. Burnett P.J. (1982) Cytoplasmic Inclusion Bodies in *E. Coli* Producing Biosynthetic Human Insulin Proteins. Science 215 687-689.

Yanisch-Peron C. Vieira J. Messing J. (1985) Improved M13 phage vectors and host strains : nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 103-19.